DOCUMENT RESUME

ED 440 293 CE 080 066

AUTHOR	Custer, Rodney L.; Schell, John; McAlister, Brian D.; Scott, John L.; Hoepfl, Marie
TITLE	Using Authentic Assessment in Vocational Education. Information Series No. 381.
INSTITUTION	ERIC Clearinghouse on Adult, Career, and Vocational Education, Columbus, OH.
SPONS AGENCY	Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE	2000-00-00
NOTE	86p.
CONTRACT	ED-99-CO-0013
AVAILABLE FROM	Center on Education and Training for Employment, 1900 Kenny Road, Columbus, OH 43210-1090 (Order No. IN381, $9.75). Tel: 800-848-4815, ext. 24277 (Toll Free); Fax: 614-292-1260; Web site: http://cete.org/products/index.html. For full text: http://www.ericacve.org/fulltext.asp.
PUB TYPE	Collected Works - General (020) -- ERIC Publications (071)
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	*Constructivism (Learning); *Education Work Relationship; Educational Assessment; Educational Research; Evaluation Methods; Learning Theories; *Performance Based Assessment; Postsecondary Education; Secondary Education; Self Evaluation (Individuals); *Student Evaluation; *Vocational Education

ABSTRACT

This guide addresses various aspects of authentic assessment as it is used in vocational education. The guide contains five chapters, each written by different authors. Following an overview and definitions of terms by Rodney L. Custer, John W. Schell discusses the theoretical foundations of authentic assessment, reviewing psychological, cognitive, and sociological views of learning. The chapter provides an extended example of an authentic assessment practice that connects authentic teaching, learning, and assessment with learning theory. In Chapter 3, Brian McAlister's literature review explores the questions of the inherent value of authentic assessment and its effectiveness in promoting learning. The chapter presents the claims made on behalf of authentic assessment and the research evidence related to those claims. In Chapter 4, John Scott details authentic assessment strategies and tools, including those that students can use to assess their own learning. In the concluding chapter, Marie Hoepfl discusses federal and state initiatives for using authentic assessment, presenting the issues, obstacles, and challenges surrounding its use on a large scale. (Contains 106 references.) (KC)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 440 293

Using Authentic Assessment in Vocational Education

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Information Series No. 381

by Rodney L. Custer
John Schell
Brian D. McAlister
John L. Scott
Marie Hoepfl

ERIC
Clearinghouse on Adult, Career, and Vocational Education

CE 080 066

BEST COPY AVAILABLE

Using Authentic Assessment in Vocational Education

Information Series No. 381

Rodney L. Custer, editor
Illinois State University

John W. Schell
University of Georgia

Brian McAlister
University of Wisconsin-Stout

John Scott
University of Georgia

Marie Hoepfl
Appalachian State University

ERIC Clearinghouse on Adult, Career, and Vocational Education
Center on Education and Training for Employment
College of Education
The Ohio State University
1900 Kenny Road
Columbus, OH 43210-1090

2000

Funding Information

Project Title:	ERIC Clearinghouse on Adult, Career, and Vocational Education
Contract Number:	ED-99-CO-0013
Act under Which Administered:	41 USC 252 (15) and P.L. 92-318
Source of Contract:	Office of Educational Research and Improvement U.S. Department of Education Washington, DC 20208
Contractor:	Center on Education and Training for Employment The Ohio State University Columbus, Ohio 43210-1090
Interim Executive Director:	W. Michael Sherman
Disclaimer:	This project has been funded at least in part with Federal funds from the U.S. Department of Education under Contract No. ED-99-CO-0013. The content of this publication does not necessarily reflect the views or policies of the U.S. Department of Education nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.
Discrimination Prohibited:	Title VI of the Civil Rights Act of 1964 states: "No person in the United States shall, on the grounds of race, color, or national origin, be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any program or activity receiving federal financial assistance." Title IX of the Education Amendments of 1971 states: "No person in the United States shall, on the basis of sex, be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any education program or activity receiving federal financial assistance." The ERIC Clearinghouse project, like every program or activity receiving financial assistance from the U.S. Department of Education, must be operated in compliance with these laws.

Contents

Foreword	v
Executive Summary	vii
Authentic Assessment–Basic Definitions and Perspectives (*Rodney L. Custer*)	1
Basic Definitions	2
Overview of the Monograph	4
Think about Authentic Learning and *Then* Authentic Assessment (*John W. Schell*)	7
The American Worker as a "Thinker"	7
Learning Transfer and Authentic Assessment	8
Psychological and Cognitive Views of Learning	9
Cognitive Approaches to Learning	10
Sociological Views of Learning	11
Authentic Instructional Strategies	14
Connecting Authentic Teaching, Learning, and Assessment with Learning Theory	16
Where Do We Go from Here?	18
The Authenticity of Authentic Assessment: What the Research Says...and Doesn't Say (*Brian McAlister*)	19
Authenticity and Authentic Assessment	19
Psychometric Issues	22
Authentic Assessment—The Claims	23
Authentic Assessment—The Research	25
Summary	29
Authentic Assessment Tools (*John Scott*)	31
Connecting, Reflecting, and Feedback	31
Self-Assessment	35
Strategies and Tools	36
Summary	46

49 **Large-Scale Authentic Assessment** (*Marie Hoepfl*)

50 Federal Mandates and Initiatives
51 Statewide Efforts to Use Authentic Assessment
53 Implementation Issues
54 Obstacles and Challenges
61 Preliminary Reactions to the Large-Scale Use of Authentic Assessment
63 Successful Models of Implementation
65 Suggestions for Implementation
66 Summary

69 **References**

Foreword

The Educational Resources Information Center Clearinghouse on Adult, Career, and Vocational Education (ERIC/ACVE) is 1 of 16 clearinghouses in a national information system that is funded by the Office of Educational Research and Improvement (OERI), U.S. Department of Education. This paper was developed to fulfill one of the functions of the clearinghouse—interpreting the literature in the ERIC database. This paper should be of interest to vocational education teachers, researchers, and graduate students.

ERIC/ACVE would like to thank the authors for their work in the preparation of this paper.

Rodney L. Custer is Department Chairperson, Department of Industrial Technology, Illinois State University. He previously taught at the University of Missouri-Columbia. He was the editor of the ***Journal of Industrial Teacher Education*** from 1996-1998. He is the recipient of the 1999 Outstanding Teacher Educator Award, Council on Technology Teacher Education; the designation as Distinguished Technology Educator, International Technology Education Association; and the Council on Technology Teacher Education's *Silvius/Wolansky Outstanding Manuscript Award* for authoring the ***Performance Standards Handbook.***

John W. Schell is Associate Professor, Department of Occupational Studies, University of Georgia. His research is focused on applications of situated cognition theory as a foundation for learning and teaching advanced vocational knowledge, skills, and attitudes. He is Co-Managing Editor of the ***Journal of Vocational Education Research*** and Assistant Editor of the ***Journal of Industrial Teacher Education***. He is the recipient of the following awards from the University of Georgia: Distinguished Service Award, Kappa Delta Epsilon; the Faculty Senate Award for Teaching Excellence (Assistant Professor Category); and Outstanding Teacher, College of Education.

Brian McAlister is Associate Professor, Department of Communication, Education and Training, University of Wisconsin-Stout, where he is involved in technology teacher education. He has also taught at the University of Maryland and Pittsburg State University. For the Council on Technology Teacher Education he chairs the Collegiate Student Association Committee, and serves on the Graduate Programs Committee. He was a reviewer for the Technology Education Standards developed by the Technology For All Americans Project, International Technology Education Association.

John L. Scott is Associate Professor, Department of Occupational Studies, University of Georgia. He has a chapter on Cognitive Achievement Evaluation in the 1993 textbook ***Improving Vocational Curriculum*** published by the Goodheart-Wilcox Company, and a chapter on Using Collaborative Team-Based Learning (Cooperative/Group Learning) in the textbook ***Education and Training***

for Work published in 1996 by Technomic Publishing Co. He served as president of the following organizations: National Association of Industrial and Technical Teachers Educators, National Association of Trade and Industrial Education, Southeastern Trade and Industrial Education Conference, and Trade and Industrial Education Association of Georgia.

Marie Hoepfl is Assistant Professor, Department of Technology, Appalachian State University, where she coordinates the technology education program. She is currently serving as associate editor of the ***Journal of Industrial Teacher Education*** and president of the North Carolina Council on Technology Teacher Education. She received the American Vocational Association/Technology Education Division Outstanding Research Award in1995. She recently completed a high school design and problem-solving curriculum guide for the state of West Virginia.

The following people are acknowledged for their critical review of the manuscript prior to publication: Brian Cobb, Professor, School of Education, Colorado State University; Ginny Felstehausen, Professor, College of Human Studies, Texas Tech University; and Morgan Lewis, Senior Research Specialist, and Cindy Zengler, Program Specialist, Center on Education and Training for Employment. Susan Imel coordinated publication development, Sandra Kerka edited and formatted the manuscript.

W. Michael Sherman
Interim Executive Director
Center on Education and
Training for Employment

Executive Summary

As educational approaches turned from behaviorist to cognitive, educators have focused on embedding instruction in real-world contexts that engage students in knowledge construction. Appropriate measures of real-world learning include authentic assessments in which students apply skills and knowledge to solving authentic problems. This monograph addresses different aspects of authentic assessment related to its use in vocational education.

Following an overview and definitions of terms by Rodney L. Custer, John W. Schell discusses the theoretical foundations of authentic assessment, reviewing psychological, cognitive, and sociological views of learning. He provides an extended example of an authentic assessment practice that connects authentic teaching, learning, and assessment with learning theory.

Next, Brian McAlister's literature review explores the questions of the inherent value of authentic assessment and its effectiveness in promoting learning. He presents the claims made on its behalf and the research evidence related to those claims. John Scott then details authentic assessment strategies and tools, including those that students can use to assess their own learning.

In the concluding chapter, Marie Hoepfl discusses federal and state initiatives for using authentic assessment, presenting the issues, obstacles, and challenges surrounding its use on a large scale.

Information on the topics in this monograph may be found in the ERIC database using the following descriptors: *Constructivism (Learning), Educational Assessment, *Evaluation Methods, *Learning Theories, Self Evaluation (Individuals), *Student Evaluation, Vocational Education, and the identifier *Authentic Assessment. Asterisks indicate terms that are particularly relevant.

Authentic Assessment–Basic Definitions and Perspectives

Rodney L. Custer
Illinois State University

As a graduate student, I vividly recall the response to the question, "So, what are the latest trends in assessment?" The question was being posed to a leading expert in vocational assessment by another professional colleague. The setting was a morning cup of coffee and my interest was piqued. The answer was immediate and simple. Authentic assessment.

A decade has come and gone since that time and much has occurred, including *A Nation at Risk*, *Goals 2000*, *SCANS* (Secretary's Commission on Achieving Necessary Skills), and more. Behaviorism has largely yielded to cognitivism, with associated interest in such things as constructivism, situated cognition, metacognition, and yes, authentic assessment.

Considerable work has been done over this past decade in the area of assessment. Around the nation, states have, with varying degrees of success, developed performance standards. In most quarters, there has been a genuine attempt to target higher-order thinking skills (e.g., critical thinking and problem solving) and to emphasize connections and synthesis over fact-based disciplinary content. Predictably, the results have been mixed, with concerns about such things as "learning the basics," confusion about content, and concerns about assessment.

At the same time, much has changed. National curriculum standards, which have been developed for many of the disciplines (e.g., science, mathematics, geography, etc.), emphasize inquiry, problem solving, critical thinking, synthesis, and authentic contexts. Changes in assessment practices have also occurred. Most states and standards efforts are promoting the use of a performance component in addition to (or in lieu of) objective-based testing. At times, this has taken the form of constructed response items; in other cases, states and school systems have experimented with incorporating more extensive performance-based activities into the assessment process.

In many respects, this decade of intensive activity has served to validate much of what has been occurring for many years in vocational education. Consider emphases such as "hands on," "lab-based," coops, and internships. For years, considerable work has been invested in identifying competencies and subsequently molding them into behavioral objectives. Although some assessment remained focused on the testing of facts, there has also been a rather natural concern for observing (watching students while they do something) and evaluating the quality of completed tasks (i.e., judging projects against established

criteria). To some considerable extent, many of the practices that have been *typical* in vocational education have emerged as *alternative* in the larger academic community.

At the same time, activity in the larger academic community is informing vocational education and the two have been drawn more closely together. Vocational education research and practice are being informed by the insights of cognitive learning theory. Those from traditional academic areas are looking to vocational educators for help with authentic contexts and activities. And both are learning more about the complex interactions and connections between authentic learning and assessment.

This monograph was conceptualized as a kind of contemporary retrospective analysis. All of the authors have, in various ways, conducted our professional work in areas that we would have a difficult time defining as either *vocational* or *academic*. Actually, it has been both. Collectively, we have worked actively and in various ways with the National Science Foundation, national and state departments of education, and the National Research Council. We have provided leadership to national standards projects and have been active with the American Educational Research Association (AERA) and the Association for Career and Technical Education (ACTE, formerly the American Vocational Association). As such, we bring a rich and varied set of experiences and perspectives to this discussion of authentic assessment in vocational education. We like it that way and believe that this mix of experiences has enriched our thinking. Throughout the pages of this monograph, we have not attempted to restrict our vision to only those materials that are most applicable to vocational, career, or technical education. Rather, we have attempted to address the key issues from within our varied and mixed perspectives. Our sense is that this mirrors the best of what is occurring across education.

Basic Definitions

Before moving into an overview of the chapters, it will first be helpful to clarify some terminology related to assessment. Three commonly used terms are alternative, authentic, and performance assessment. Conceptually and in practice, these terms tend to describe similar things.

Alternative Assessment

Perhaps the least descriptive and useful is the term "alternative assessment." As the term indicates, alternative assessments are essentially any assessment practices or tools that are different from traditional practice; more specifically, different from *paper-and-pencil tests*. A more informative approach is that taken by Neill (1997), associate director of the National Center for Fair and Open Testing. Neill has identified seven defining principles for *new* assessments developed by the National Forum on Assessment. These principles have received widespread support among educators and civil rights leaders, based on a desire for radical

reconstruction of assessment practices as well as an emphasis on student learning as central to assessment reform. The seven principles endorsed by the forum are as follows:

1. The primary purpose of assessment is to improve student learning.
2. Assessment for other purposes supports student learning.
3. Assessment systems are fair to all students.
4. Professional collaboration and development support assessment.
5. The broad community participates in assessment development.
6. Communication about assessment is regular and clear.
7. Assessment systems are regularly reviewed and improved.

Actually, there are many different definitions offered for alternative assessment and no single definition prevails. According to Hamayan (1995), alternative assessment refers to procedures and techniques that can be used within the context of instruction and can be easily incorporated into the daily activities of the school or classroom. Huerta-Macias (1995) contrasts alternative assessments with traditional testing by placing the emphasis on integrating and producing rather than on recalling and reproducing. These authors also note that the main goal of alternative assessments is to gather evidence about how students are approaching, processing, and completing real-life tasks in a particular domain.

The term alternative assessment provides an umbrella for a variety of nontraditional assessment methods and techniques such as direct assessment, authentic assessment, and performance assessment (Butts 1997). However, given the growth and refinement that have occurred over the past decade, the term suffers from a lack of precision.

Authentic Assessment

Authentic assessments are essentially those that embed assessment in real-world contexts. Wiggins (1993) describes authentic assessment as tasks and procedures in which students are engaged in applying skills and knowledge to solve "real-world" problems, giving the tasks a sense of authenticity. He goes on to define authenticity as that which replicates the challenges and standards of performance typically facing writers, businesspeople, scientists, community leaders, designers, and technical workers. To design an authentic assessment activity, teachers must first decide what are the actual performances that they want students to be good at and then they must decide how they can frame learning experiences in a meaningful context that provides the connections between real world experiences and school-based ideas (Lund 1997).

A number of criteria have been used to define and describe authentic assessment. Among these are the following (Lund 1997; Wiggins 1993):

- Engaging and worthy problems or questions of importance to students,
- Replicas of or analogies to the kinds of problems faced by adult citizens and consumers or professionals in the field,

- Tasks that require the student to produce a high-quality product and/or performance,
- Transparent or demystified criteria or standards,
- Response-contingent challenges in which the effect of both process and product/performance determines the quality of the results,
- Emphasis on "higher-level" thinking and more complex learning,
- Evaluation of the essentials of performance against well-articulated performance standards often expressed as rubrics, and
- Assessments so firmly embedded in the curriculum that they are practically indistinguishable from instruction.

At a minimum, authentic assessments are those that require real-world applications of skills and knowledge that have meaning beyond the assessment activity (Archbald and Newmann 1988). Howevèr, a review of the criteria listed here shows that the concept also has been extended to include complex performances, creation of significant products, and accomplishment of complex tasks using higher-order cognitive skills.

Performance Assessment or Performance-Based Assessment

At the most basic level, performance assessment involves asking students to do something and then observing and rating the process and the finished product against predetermined criteria or a standard. As with other terms used to describe the various forms of assessment, other definitions of performance assessment tend to blur this distinctive meaning. For example, Herman (1999), associate director of the National Center for Research on Evaluation Standards and Student Testing, states that the "essence of performance assessments–whether in the form of open-ended questions, essays, experiments or portfolios–is that they ask students to create something of meaning" (online, n.p.). Herman continues by observing that good performance assessment involves complex thinking and/or problem solving, addresses important disciplinary content, invokes authentic or real-world applications, and uses tasks that are instructionally meaningful. Stated in this way, performance assessment sounds very much like authentic assessment.

In reality, the distinctions among terms are probably relatively small and probably insignificant. For our purposes in this monograph, we have chosen to use the term *authentic assessment*, since it tends to draw the boundary more broadly than performance assessment (authentic assessment typically involves some form of performance) and more precisely than alternative assessment (which typically includes everything but traditional testing).

Overview of the Monograph

The four chapters that comprise this work address distinctively different aspects of authentic assessment. In chapter one, John Schell discusses the theoretical underpinnings of authentic assessment. Whereas vocational education has a long history of behaviorist-oriented, competency-based education, authentic assessment has increasingly been informed by contemporary cognitive and sociological

learning theory. An important focus of the chapter is on the value of authentic learning and assessment practices as a mechanism for promoting learning transfer. In the second chapter, Brian McAlister provides a review and synthesis of what the research literature has to say about the value of authentic assessment. This "value question" has two important dimensions. First, the question is asked about the inherent value of authentic assessment as an approach to assessment. The second question has to do with the effectiveness of authentic assessment as a mechanism for enhancing and promoting student learning. Chapter three moves to the more pragmatic end of the continuum. After an initial discussion of three key concepts associated with authentic assessment (connecting, reflecting, and feedback), John Scott provides a comprehensive overview of the "tools" that are commonly used for authentic assessment. In the final chapter, Marie Hoepfl addresses one of the more perplexing issues associated with authentic assessment: the issues and challenges of using authentic practices for large-scale, high-stakes assessments.

We have enjoyed the discussions that led to the development of this monograph. We hope that you will enjoy it and that it will serve to extend your thinking about the nature of assessment in general and authentic assessment in particular.

Think about Authentic Learning and Then Authentic Assessment

John W. Schell
The University of Georgia

How often have educators heard, or been asked, the question, "Where am I ever going to use this stuff?" At other times, the question is posed less directly in the form of student behavior, such as apathy, open resistance, cramming for tests, or simply "going through the motions."

With this in mind, one of the many things educators can do is to engage students in topics and issues that are real, meaningful, and engaging. Although the focus of this monograph is on authentic assessment, it is important to begin the discussion with authentic learning. Consequently, most of this chapter focuses on what we think we know about learning. It will do little or no good, and may even do harm, to adopt new assessment practices without proper alignment between approaches to instruction (with its underlying assumptions about learning) and new ways of thinking about assessment.

The American Worker as a "Thinker"

In addition to the plea for authentic learning experiences as a base for authentic assessment, the call for authenticity is being heard from another sector; namely, from employers, who are looking for a new type of worker. Many argue that today's worker should be both a "thinker" and a "problem solver." This concern was identified in the 1991 SCANS report, which indicated that expert workers will be unable simply to pick up these competencies haphazardly. The teachers of future generations must engage students in more demanding school activities designed to promote the development of higher-order thinking and problem solving. Parnell (1995) made this case in support of what he called "contextualized learning." These points suggest a major reform of school curricula and methods of assessment. Parenthetically, Schell and Rojewski (1995) have argued that higher-order thinking skills should extend to teachers as well, since they are uniquely positioned to model the use of these skills to students.

The profession is gradually realizing that Parnell is correct. Learning advanced thinking skills occurs best when it starts in school and continues throughout life. Yet, the traditional fact-based curriculum and subsequent "brain dump" assessment does little to prepare future generations to function as thinkers, problem solvers, and lifelong learners. Many experts believe that today's fact-based curriculum requires a level of "learning transfer" that extends far beyond what could reasonably be expected of most students.

Think about (Schell)

Learning Transfer and Authentic Assessment

What has prompted the current interest in authentic assessment? Is it just the next educational fad or does it represent something more substantive? One useful approach to assessing the value of authentic assessment and learning is to begin with a discussion of learning transfer. Many theorists have come to believe that learning is more mobile when the contexts for learning and application are similar (Lave 1988). This view calls into question some assumptions traditionally made about how learning occurs and is later used. Essentially, the question has to do with the relationship among teaching, learning, content, and context, as well as the resulting impact on learning transfer. A closer examination of these complex relationships is in order.

Traditionally, an implicit assumption of educators has been that classroom learning will more or less be transferred to other problems encountered at work, at home, or in other classroom settings. This "transfer assumption" is so pervasive that many have come to believe that it is a routine and predictable artifact of teaching and learning. In fact, this belief is the heart of the prerequisite curriculum so common at almost all levels of the U.S. educational system. Curriculum designers often assume that arithmetic learned in a basic math class will transfer as students encounter algebra in a subsequent class. This principle is customarily represented in curricula ranging across the entire educational spectrum from the elementary school to the top research universities. Assuming for the moment that this assumption is true, it would make sense to require a basic math course prior to advanced applications such as algebra or chemistry. Unfortunately many researchers now argue that "transfer is very difficult to obtain" (Detterman 1993, p. 7). It is probably not a routine and predictable learning event as much of the educational community has presumed.

It is helpful to preface the remaining discussion with an operational definition of learning transfer. From a psychological perspective, transfer is defined as the degree to which a behavior will be repeated in a new situation (Detterman 1993). Distinguishing between "near" and "far" transfer and constructs such as "surface structure and deep structure" further refines the concept. Near transfer is knowledge learned and used in similar situations. Far transfer is thought to occur when knowledge is applied in a context dissimilar from the one in which it was learned. Typically, far transfer is the desired goal of the learning transfer process. Learning transfer is a little like hitting the educational home run. It is effective, efficient, dramatic—and rare. However, hitting singles and doubles can more predictably score runs. The same is true when thinking about teaching strategies that promote transfer. If we can teach in contexts similar to how the information will be used then we have a better chance for multiple uses of information. This is the principal argument for authentic teaching, learning, and (later) assessment. Transfer is more likely to occur (even if it is near transfer) when instructional and application settings are nearly identical (ibid.).

In spite of the fact that transfer is less than routine and predictable, it makes sense to enable multiple uses of learned information through a variety of teaching

strategies. It is also useful to examine a variety of learning theories for what they have to say about learning transfer. The following section is a brief "sampler" of the various educational and psychological theories that have historically informed educators' views on learning. After a brief review of these approaches, the discussion turns to the related area of cognitive science, which provides a base for "constructivism." As part of this discussion, we examine the meaning of "mental frameworks" and "associations." Finally, we discuss sociological learning theories and how they are thought to support authentic instruction and assessment.

Psychological and Cognitive Views of Learning

The basic frameworks for psychological learning theories have many storied historical and traditional roots. In fact, much of the foundational thinking in the delivery of vocational education can be traced to John Locke's notion that the mind of the learner can be represented as a "tabula rasa" or a "blank slate." However, research has shown that individuals are endowed with suspended "biologically preformed abilities" that may lie dormant until awakened by the input of appropriate data (Phillips and Soltis 1998, p. 13). For example, speech may be a latent ability that is enabled only by a child hearing spoken language. Although many modern psychologists do not agree with Locke's explanation of preformed abilities, many of the traditional and modern theories of learning rely on "mental frameworks" or learning by "association."

Behavioral Approaches to Learning

In traditional vocational education, psychological learning theories have been used to focus on education and training for specific jobs or skill sets. Behavioral researchers such as Hull, Thorndike, and Skinner are the primary proponents of these adopted theories. The research that supports behaviorism comes from careful scientific study of animal behavior. Researchers believed that inferences could be made with regard to human behavior because of the biological similarities between man and lower animals (Phillips and Soltis 1998).

Early behaviorists were not particularly concerned with how individuals acquired new knowledge or the origins of these ideas. They were more concerned with how individuals acquire new behaviors. Behavioral psychologists are concerned with two general areas, classical and operant conditioning. Both are built on stimulus (S)–response (R) associations. Classical conditioning involves an associated or "conditioned" response, which later substitutes for the original stimulus. Pavlov's work with dogs is an example: Food (S) is presented and the dog salivates as a response (R) (Watson 1930). Later, a bell is rung with the presentation of food. Ultimately, the bell can be shown to replace the food as the stimulus causing the dog to salivate. In classical conditioning, a stimulus is presented and the animal exhibits some type of behavior and then receives a reward for its performance (Thorndike 1913). Skinner (1966) later determined that reinforcement does not need to be presented with every successful performance. He found that "he could 'shape' the behavior of his laboratory animals in startling ways just by the judicious use of rewards" (Phillips and Soltis 1998, p. 28).

Think about (Schell)

E.L. Thorndike extended behaviorism in his work with operant conditioning. He also used research animals—in this case, a cat in a box with a release mechanism that, when operated, would open the cage door or produce food. Thorndike recorded the cat's progress over successive trials. This documentation has become commonly known as the learning curve (ibid.). Over a number of trials, the cat gradually got the idea through successive approximations. This led to Thorndike's laws of exercise and effect. The law of exercise holds that the more a stimulus and response connection is activated the stronger it becomes. The law of effect addresses the pleasure that one gets from successful learning, thus increasing the probability of future attempts.

Elements of behaviorism can be found in today's practice of academic and vocational education. For example, operant conditioning is the theoretical basis of behavioral modification in which teachers provide systematic rewards for appropriate classroom behaviors. Many other teachers "manage" the behaviors of students through a "token economy" in which rewards for privileges are provided to those who exhibit desired responses. Elements of behaviorism are also present in competency-based instruction (CBI). In a well-designed CBI system, tasks are identified through task analysis and are presented to the learner in the form of performances (or behaviors) to be mastered in requisite order (Mager 1975). This linear presentation of competencies is based on the assumption of routine and predicable transfer. First, basic information must be acquired before more advanced applications are possible. Other points of view on this topic are discussed later.

Cognitive Approaches to Learning

In recent years, vocational and academic instruction have drifted back toward the future. We are revisiting some of the theories on mental frameworks that date as far back as John Locke's Atomistic Theories and, more recently, Piaget's developmental theories (Phillips and Soltis 1998). Yet, these theories are also futuristic as radical constructivists such as von Glasersfeld (1995) extended the work of Piaget and of Bruner (1966). Earlier constructivists viewed learning as active engagement through which new ideas are "constructed" based on the current or past knowledge of the learner. Schema or mental models were thought to provide cognitive structures for the extension of present knowledge and the creation of new understanding.

Piaget, a biologist by training, suggested that, as children progress through "stages of development," they acquire new capabilities through adaptation, assimilation, and accommodation. In his sensorimotor stage (ages 0-2), Piaget suggested that the development and refinement of physical movement shape and drive behaviors. In the preoperational stage (ages 3-7), only physical objects and their manipulation are represented in the developing mental frameworks. In the concrete operational stage (ages 7-11), certain logical structures are constructed from physical encounters (Phillips and Soltis 1998). Here, abstract concepts of the mind are increasingly possible, but are mostly generated through the physical

manipulation of objects. In the last stage (formal operations), adults are able to solve abstract problems using various levels of reasoning (ibid.).

More radical constructivists seek increasingly to build upon complex mental structures, but also require individuals to cope with and interpret experiences (von Glasersfeld 1995). Vygotsky, a Soviet-era Russian psychologist, little known in this country until recent years, argued for the importance of social influences on learning. Vygotsky's research differs from Piaget by suggesting that chronological conceptions of development might be replaced by "zones of proximal [or potential] development...'ZPD'" (Phillips and Soltis 1998, p. 59). This allows for children to develop at different rates, but certainly not according to stages roughly organized by age.

John Dewey was known as both a philosopher and a learning theorist in his extensive and productive career. His belief in the importance of experience and the use of logic for the purpose of solving problems makes him a candidate for extending principles of constructivism even further into the world of social influences (McDermott 1981). Dewey noted that "purposeful learning in social settings [is] the key to genuine learning" (Phillips and Soltis 1998, p. 56.). In this way Dewey's beliefs were compatible with the constructivist movement and also with the emerging social views of learning, which are explored later.

Constructivism and Authentic Learning/Teaching

These more recent cognitive explanations of learning provide a context for understanding how and why "authentic" instructional and assessment strategies promote the teaching of critical thinking and problem solving. Using constructivism, a teacher can design purposeful educational activities that require learners to build on and extend their mental models.

Constructivist views free curriculum designers from the linear assumption of focusing on "basics first" as the primary strategy for promoting learning. The instructional design process is expanded to explore a "global view" before focusing in on "local" details (Brown, Collins, and Duguid 1989; Schell and Rojewski 1995). The teacher provides a roadmap of the entire subject to be learned while allowing students to construct their understanding of the topic. Learners assume increasingly more control over the sequence in which they want to engage their learning and are free to explore the various local details of the topic. They can build their own mental frameworks in ways that are natural to them, unencumbered by a superimposed logical sequence.

Sociological Views of Learning

In recent years, Jean Lave and Etienne Wenger have written about learning in a different way, from a very different perspective. As ethnographers, they have employed principles of sociology, while emphasizing the importance of context and participation in communities of practice as critical of elements in the learning

process. These ideas are potentially important to the design, delivery, and assessment of both context-based vocational and academic education.

Many of the ideas that are expressed in this section will be familiar to many readers. The careful scholar will notice many similarities between these ideas and those discussed over the past 150 years by such writers as James, Dewey, and Vygotsky. The difference here is that these more recent contributions have come from scholars outside traditional cognitive science.

Jean Lave, an anthropologist, has studied learning as it occurs in natural settings. Her research often examines elements of partial and full membership in some type of community. This body of literature has come to be popularly known as situated cognition, or legitimate peripheral participation. Lave's collaborator, Etienne Wenger, has extended this body of research in his most recent publication Communities of Practice (1998). Like the research on constructivism, this research also has important possibilities as a framework for authentic instruction and assessment.

Situated Cognition

From her naturalistic studies, Lave coined the term "situated cognition" to describe the cognitive process as a "nexus of relations between the mind at work and the world in which it works" (Lave 1988, p. 1). She further proposed that cognition is not just a psychological phenomenon, but rather "stretched across mind, body, activity and setting" (ibid., p. 18). This view of cognition is not new, but rather lends increased credence to the foundational work of educational theorists and philosophers such as Dewey (1974) and Vygotsky (1978) who wrote about social learning and the importance of instructional context.

Lave and others have researched learning in everyday life contexts, as opposed to abstract classroom or laboratory conditions (Lave 1988; Lave and Wenger 1991; Resnick 1987). They have found that when individuals address problems requiring the same knowledge, the context in which the person was engaged greatly influenced how they used information to solve a problem. Lave and Wenger (1991) give an example of individuals attempting to follow weight reduction diets. In their own kitchens, dieters relied on estimation techniques, often physically dividing food into appropriate portions. However, in a classroom setting these same dieters attempted to use paper-and-pencil approaches to dividing fractions. This and other research indicates the importance of learning contexts to how problems are thought about as well as how solutions are generated.

In reporting their research, Lave and Wenger (1991) used the term "legitimate peripheral participation" to describe how individuals gain opportunities to use their learning as a member of a community. In this community role, individuals must make a legitimate contribution to a situation that is valued and considered "authentic" by the learner. These contributions initially are likely to be at edges (or the periphery) of the socially constructed community. As new members progressively demonstrate competence, other members of the community gradually

allow novices to engage in more complex activities. In this way, learners are eventually affirmed as full-fledged members. Through participation, learners also construct their identity relative to the community. As a result, learners achieve a mental "meaningfulness" that comes from participation as members of a valued community.

Communities of Practice

Wenger (1998) has extended this work into a more formalized construct, termed "Communities of Practice." Learning is viewed as a central element that connects the interaction of meaning, practice, community, and identity. Meaning is a way that we use our increasing abilities to create meaning from our lives and our work. Practice (or collective participation) is a way in which our community constructs a mutual history, collective social resources, and common ways of looking at the world. These commonly held values guide our actions and promote continued engagement in the business of the community. Community consists of the social networks, which define our enterprises as worth pursuing and recognizes the work of individuals as competent. Identity is a way of talking about how learners change as they learn. In this way, learners create personal histories of how they have become members of a community of practice.

Based on the principles of community of practice, it is the "meaningfulness" that a learner attaches to the content that makes multiple uses of information possible. These writers do not acknowledge learning transfer as a construct. Rather, they believe that learning is a new event in each situation. Wenger believes that community members ultimately achieve such meaning through the interaction of their participation and the reification of imaginary and real objects that represent the values of the community. For example, schoolteachers have a number of imaginary symbols that represent their own communities of practice. They might be intangibles such as the common beliefs held among our colleagues with regard to discipline in the classroom. These beliefs can also be actualized for faculty and students in the form of a handbook. It is participation as teachers in the valued enterprise of educating youth, and the associated real and imaginary symbols that give a professional community its meaningfulness. This also represents the meaningfulness that shapes our professional identities (Wenger 1998).

Learning within communities of practice is thought to have several characteristics, including the following (Wenger 1998):

- The ability to negotiate new meanings—teach for meaning, not for mechanical recall of isolated information.
- Creating new mental structures—teach with enough structure and continuity to promote meaningful new mental models while reconsidering prior learning that might be inappropriate.
- Learning as both experiential and social—teach in realistic social settings that require the learner to engage deeply with the community.

Think about (Schell)

- Learning as a matter of engagement—teach using strategies that require learners to engage with material that they find interesting. This can be an instructional springboard to introduce learners to ideas and concepts that they might not initially see as inherently interesting.
- Learning as an agent of change— teach with the knowledge that we are what we learn. Allow learners to change their positions. This can be done through articulation and reflection strategies.

Authentic Instructional Strategies

Constructivist and situated cognitive research has important implications for teaching and learning. Teachers who place high value on learning in authentic contexts usually organize their instructional day very differently. One of the most obvious differences is devoting less time to describing content, with more time spent on enabling students to "experience" the use of the information in real or realistic settings. Thus, context and social relationships become important instructional considerations and frameworks.

When designing instructional and assessment activities, it is important to ask, When is "real" real enough? Is it authentic enough when we employ a computer or role-playing simulation? The answers are complex, which usually means both "yes" and "no" are correct answers. The problem is further exacerbated by the fact that the answer is often individualized to the learner. Both the constructivist and the situated cognition teacher would agree that the context must be realistic enough to the learner to build on existing mental schema or to engender meaning through participation and a deeper understanding of the community while it is in action (Wenger 1998). The key is to find strategies to engage learners in community activities that capture their imaginations.

Authentic Learning, Teaching, and Assessment—An Example

Teachers wanting to implement a program of authentic instruction and assessment must consider several key points. First they must pay more attention to the important roles played by physical and social contexts of learning. Second, viewing learners as members of a community of learners raises issues of relationships, identity, trust, and power (Schell and Black 1997; Wenger 1998). This type of teaching requires teachers to be flexible, alternating between direct and facilitated instruction as appropriate and desirable.

The following example illustrates many of the instructional activities that could support an authentic learning experience leading up to authentic assessment. This example may be more complex than those that would be implemented in a single classroom. Not every step described here will be required with every student; this example illustrates a comprehensive range of authentic learning and assessment procedures.

As a high school technology education teacher, you require your students to serve as an intern in a field related to their current career plan. Heather is a first-semester senior who is considering a career in engineering. Currently, she is planning to attend a technical institute next fall where she will work toward an associate degree in pre-engineering. However, these plans could change depending on how well she does in the technical school and how her finances play out. Heather has been an average to good high school student, but in your professional opinion, she is capable of much more. The internship that you have arranged for her is with the Johnson-Brown Company, a civil engineering firm that has just received a contract to design a new bridge in a rural area of China. Your contact at Johnson-Brown has agreed to allow Heather to become a novice member of the team on the project. Although she is far from being an engineer, you ask your contact, Ms. Patty Freeman-Young, to give Heather meaningful work to do on the project. Patty agrees, telling you that the project is their first for a Province in China and they need lots of background information. Heather's first job will be to conduct background research using the Internet and contacts at the Chinese Consulate. Heather's specific assignment will be to research some of these considerations and prepare a brief that will inform the project engineers as they create the bridge design.

As a real member of the team, with real and important work to do, Heather will experience the culture within the Johnson-Brown Company as well as the daily practices of engineers. She will directly observe how principles of physics are applied to an authentic problem. In addition, she will learn a great deal about life in rural China and will have the opportunity to explore aspects of Chinese construction technology, materials, and practices, which must be reflected in the Johnson-Brown design. As a result, Heather will be exposed to problems that are routinely encountered and solved while considering the balance of the Chinese culture, public safety, and investment in infrastructure.

Most of Heather's internship goes very smoothly. She proves to be highly energized by her work, exclaiming "I am doing work that has a real purpose! I really like doing this type of work." Yet, minor problems emerge as some of the engineers find Heather's questions and enthusiasm rather distracting. Patty Freeman-Young calls at home one night with an idea that might make life easier for Heather. She requests that you come to the Johnson-Brown facility to explain the purposes and educational advantages of the internship to the engineers. Although this may not completely resolve the issue, the approach sensitizes the company to its responsibility toward younger workers.

The approach proves to be helpful, not only to Heather, but to the staff engineers at Johnson-Brown. You anticipate this problem with future

interns and prepare a 5-minute PowerPoint presentation that can be given before the next student internship begins. The point to be made here is that relying on professionals outside the school district often requires a little preservice preparation for those who will be interacting with the student-learner.

As thought provoking and challenging as this activity might be for the right student, it is not yet a complete authentic learning experience. Authentic educational experiences must also be "examined" experiences. It is not enough for a learner just to have community experiences. The meaning of that experience in their lives must be probed and used purposefully. An important part of the collaboration between you and the Johnson-Brown Company is the opportunity that you provide for Heather and her classmates to reflect on the meaningfulness of their internships. This can be done through the use of class time set aside for learners to articulate what they have learned and to reflect on its meaningfulness to their personal plans for the future. Heather might be asked to discuss some of the new things she has learned from her research on the Chinese bridge project. As her teacher, you might want to ask her to engage in some thinking about her own thinking. The technical term for this activity is metacognition. You might ask probing questions such as "How did you learn that?" Or, "What thought processes did you use to come to that conclusion?" At first, Heather is caught off guard by these types of questions. But, she soon begins to anticipate these kinds of reflective questions and incorporates them into her reflections. The importance of metacognition is that it (1) clarifies for Heather her mental and social approaches to solving problems, (2) provides examples of her problem solving for other students as they are challenged to think about their own thinking, and (3) provides the teacher with instructional moments where assistance can be provided when it is needed. Reflection strategies also provide opportunities for instruction in which students learn that information learned in internship experiences can have multiple uses. In other words, learning begins to transfer from one context to another. Psychologically speaking, Sternberg and Frensch (1993) observe that teachers can promote transfer through direct and overt actions, expecting and requiring learners to use information to solve a variety of problems.

Under careful supervision, Heather's internship experiences can also encourage her to see how social contexts enable learning. She will likely have a much more highly developed cognitive framework with regard to the work of a civil engineer (von Glasersfeld 1995). As a result of the internship, Heather could be more engaged with her technology education schoolwork, find new meanings from her experience, and be changed as a person because of her learning (Wenger 1998).

These strategies open the door for authentic assessment. In fact, a reflection period (such as the one described here) can be considered a

form of authentic assessment. Other assessment strategies that could be included in such an activity are reflective journals, portfolios, a video documentary, or even a detailed research paper. Heather was required to document her experience using a portfolio approach. Hers included (1) a statement of purpose, (2) seven short reflection papers that described important events, (3) evaluation reports from Patty Freeman-Young, and (4) examples of the work that she performed while in her internship. She interspersed many photos in the paper copy of her portfolio. In addition, Heather was allowed to use space on the Johnson-Brown server computer to create a webpage where she stored her portfolio in electronic format.

Connecting Authentic Teaching, Learning, and Assessment with Learning Theory

Heather's internship at Johnson-Brown Civil Engineering is based on both psychological (constructivist) and sociological (communities of practice—situated cognition) learning theories. The combination of these theories provides an opportunity to create educational opportunities that deeply engage students with meaningful work and could even cause them to be "turned on" by learning. Heather's internship is connected with these theories in the following ways.

Heather is learning in a situated context. The use of realistic settings has great implications for learning and the later use of acquired information. Some teachers might be tempted to substitute a simulation or a computerized approach, thinking that it will also contextualize learning. A simulation might work for some learners and even be easier on the teacher. Whatever the approach taken to address this problem, a general rule could be helpful: Make the learning setting as realistic as possible. This will increase the probability of "meaning making" among learners. When it is at all practical, get the students out of the classroom, off the campus. Require them to interact with the world as a member of a learning community.

Heather is participating in a community of practice. Because Heather's teacher took the time to insist on meaningful work for her to do, Heather had a greater opportunity to become a member of the Johnson-Brown community. Her report on Chinese culture and building practices might prove to be very helpful in the project. It could also save someone else a great deal of time researching the information. The benefit for Heather is that through her participation in the community she now understands much more about the work and daily life of a civil engineer. She will now have mental images and frameworks that will help her understand the pre-engineering curriculum at the technical college. Potentially, this authentic experience will give her course of study much more relevance (Wenger 1998).

Heather is "constructing" her knowledge at work and at school. Because of the way that the internship was set up, Heather was required to articulate and reflect

Think about (Schell)

on her new knowledge. Through the act of reflection, she was forced to interpret her experiences thoughtfully and make inferences for their meaning to her life and her future. Consistent with constructivist theories, Heather's new knowledge is enriching her old and extending her mental images (Bruner 1966). This can be an emotional experience for some learners. By constructing her knowledge, she may well develop a passion for her work that she has never before experienced. In this way learning will change her (Wenger 1998).

Heather is also constructing her knowledge in another way. As she shares her experiences and listens to others in the learning community, other students and their teacher are interpreting the meanings of individual experiences for the entire group. This is also a form of radical constructivism in which the entire community participates and benefits from one another (von Glasersfeld 1995).

Heather is examining her own learning at it occurs. Because you had the foresight to ask her to learn about her own learning, Heather may have had insights into how she learns and subsequently uses information to solve problems. Such information can be very helpful as she learns to control and direct her knowledge (Brown et al. 1989; Schell and Rojewski 1995).

Where Do We Go From Here?

The suggestions made in this chapter have major implications for many schools. This is especially true now that many schools are adopting block scheduling and additional time and resources can thus be devoted to off-campus experiences. However, there are many practical, logistical, and political reasons why schools are limited in placing learners in realistic contexts such as the Johnson-Brown engineering firm. Even if politics and/or resources prevent authentic instruction, the use of simulations and role-playing can be substituted. Whatever approach is used to promote authentic teaching and learning, it is important to remember this: The more authentic it feels to the learner, the better the results and the associated transfer of learning are likely to be.

Authentic teaching and learning makes authentic assessment possible. The next chapter examines more specific assessment strategies. Experts describe in detail innovative and imaginative assessment strategies.

The Authenticity of Authentic Assessment: What the Research Says...Or Doesn't Say

Brian McAlister
University of Wisconsin-Stout

The purpose of this chapter is to report what the research says about authentic assessment. First, the claims that have been made about the benefits of authentic assessment as a mechanism for measuring student performance are discussed. Next, the claims that have been made about the benefits of authentic assessment as a mechanism for facilitating learning are examined, followed by a review of research related to authentic assessment. The chapter concludes with a brief discussion of some key issues of concern related to research and practice.

Much has been written about the promise of authentic assessment. A primary focus of much of what has been written is to promote the use of authentic assessment as a superior alternative to other forms of assessment. Although much of this work has been positive and while many of the benefits make sense intuitively, the question remains: What does the research say about the benefits and problems associated with authentic assessment? To understand some of these claims, it is important to understand the conceptual foundations of authentic assessment.

Authenticity and Authentic Assessment

The overarching theme of authentic assessment is, as the term indicates, authenticity. This thrust relates both to the authenticity of the *learning* activity as well as the authenticity of the *assessment*. One concern that is voiced throughout the literature has to do with what makes an activity "authentic." In vocational education circles, with the rich history of laboratory-based learning, this concern is much less problematic than in the more traditional academic areas. As performance and authentic assessment have moved more broadly into the academic arena and as vocational and academic education have attempted to work more closely together, the issues of "authenticity" have become more important.

Messick (1992) captures this sense, indicating that "a fundamental ambiguity pervades authentic educational assessments, namely, authentic to what?" (p. 27). He poses the question of whether assessments should be authentic reflections of classroom work or authentic reflections of the "real world." This is a subtle, but important distinction. What is meant by the real world? Sometimes students are taught using conventions that have been found to be effective and efficient

methods of teaching certain skills or concepts. Does the fact that educators commonly use various algorithms when teaching mathematics make this approach authentic? Is authenticity defined by the boundaries of the classroom? To what extent do (and should) the real world experiences of students coincide with what occurs in classrooms and laboratories? Students' perceptions of the real world may indeed be very different than those of their teachers. One could assume that authenticity means teaching in context or contextual learning. But then we are left to ascertain which of these contexts are worthy of the distinction of being considered "real" or "authentic." For example, some vocational schools have established automotive service programs that operate like service centers in automotive dealerships. Customers schedule their maintenance with students, who order parts, repair the automobiles, and when the work is completed, bill the customers. Because this is an educational experience, there are times when an instructor must intervene on behalf of the customer. Students cannot be allowed to make serious mistakes that could result in a dangerous automobile being released to the customer. This scenario poses some serious questions. If authentic assessments should reflect the "real world," How real is "real"?, Whose reality should it reflect?, and What degree of authenticity is "authentic"? It is obvious from this example that educators must apply reasonable limits on authenticity as a function of concerns such as safety, confidentiality, and more.

Others have attempted to clarify these issues by suggesting criteria to gauge the authenticity of an activity or learning experience. Newmann and Wehlage (1993) suggest that, in order for instruction to be considered "authentic," students must construct meaning and produce knowledge, use disciplined inquiry to construct meaning, and aim their work toward production of discourse, products, and performances to a level of value or meaning beyond success in school.

In order to meet these criteria, Newmann and Wehlage offer five standards (or criteria) that can be used to distinguish levels of authenticity of a learning activity:

1. To what extent are students required to use higher-order thinking skills?
2. What is the depth of student knowledge and understanding that is attained?
3. At what level does a learning or assessment activity have value and meaning beyond the classroom?
4. To what extent are students required to discuss, learn, and understand the substance of a subject?
5. How well does an assessment measure the expectations, respect, and extent of inclusion of all students in the learning process?

Newmann and Wehlage's criteria are useful because they refine and clarify the distinctions that should be made relative to the meaning of authenticity. The criteria also extend authenticity beyond simple participation in "real" experiences to active reflection on the meaning of those experiences.

From another perspective, Cronin (1993) and Tanner (1997) suggest that the concept of authenticity is relative and exists along a continuum. An example of

this would be to compare activities that might occur in a teacher education program. It makes sense that demonstrating how to use a cooperative learning technique during a microteaching activity could be considered more authentic than simply writing a paper about cooperative learning. On the other hand, using cooperative learning techniques in a class while student teaching would be considered more authentic than using the same techniques during microteaching. Drawing from this example, it is apparent that learning activities can be placed along an authenticity continuum. Cronin (1993) supports this approach by suggesting that learning activities are "neither completely authentic nor divorced from reality" (p. 78). He further suggests that our goal as educators should be to move instruction toward the more authentic end of this continuum.

Another key aspect of authenticity in assessment relates to the strategy or system that is used. In order for assessment to be considered authentic, there should be consistency between the assessment and the real-world application for which the learner is being prepared (Tanner 1997). For example, if students are expected to be able to troubleshoot the electrical system of an automobile, then the assessment strategy should be designed in such as way as to be able to tell whether they have the knowledge and skill to perform that kind of activity.

Messick (1992, 1996) has analyzed the appropriateness of using authenticity as a standard for validity in assessment. He frames the issues in terms of representation, directness, and relevance. An assessment that suffers from construct underrepresentation variance fails to test a construct adequately, because a major aspect of the construct extends beyond the measure. For example, an assessment could be designed to measure whether a student can service automobile braking systems. If, however, students are only tested on one type of braking system (e.g., disk brakes), then this assessment would suffer from construct underrepresentation. "The measurement concern of authenticity is that nothing important has been left out of the assessment of the focal construct" (Messick 1996, p. 16). An assessment that suffers from construct irrelevant variance includes information that is irrelevant to the construct being tested. For example, the purpose of the assessment could be to determine whether students can apply appropriate design principles when designing a visual message. If the assessment method is restricted to identifying the parts of a camera, the assessment would suffer from construct irrelevant variance. Thus, an assessment is considered *representative* when it is broad enough to assess adequately the constructs being tested and *direct* when it is narrow enough to not be confounded with irrelevant information. Wiggins (1993) summarized a similar point, indicating that "tests are simplified of contextual 'noise' and 'surround' to make scores more reliable. Yet we need to maximize the fidelity and comprehensiveness of the simulation for validity reasons" (p. 230).

Tanner (1997) provides a good summary of the interrelationship between authenticity and learning experiences noting that—

> [Authentic assessment] presumes that students will produce something that reflects not a narrow, compartmentalized repetition of what was presented to them, but an integrated scholarship which connects their

learning housed in other disciplines and which is presented in a setting consistent with that in which the learning is likely to be most useful in the future. (p. 14)

Psychometric Issues

Some disagreement exists in the literature regarding what sort of standards should be used to gauge authentic assessment from a psychometric perspective. Hipps (1993) argues that the assumptions underlying authentic assessment have their basis in constructivist theory. These assumptions, and associated psychometric considerations, are different from those commonly associated with traditional measurement theory. Therefore, he calls for a new set of standards that he suggests should start with trustworthiness and authenticity to replace the traditional standards such as reliability, validity, and objectivity, which are used in positivistic, quantitative research.

Reckase (1997) counters that this call for a different theoretical framework makes sense "if performance assessments are used solely as instructional tasks" (p. 12). However, if the issue is assessment then some statistical requirements are needed. He goes on to argue that "reasonable statistical requirements for sound performance assessments can be described based on current experience in the areas of (a) rater reliability, (b) test reliability, (c) generalizability, and (d) validity" (p. 3).

In a similar vein, Messick (1992) argues that in authentic assessment "different psychometric models might be employed . . . but such basic assessment issues as validity, reliability, comparability, and fairness still need to be uniformly addressed" (p. 7). He argues that "the interpretation and use of performance assessment . . . should be validated in terms of content, substantive, structural, external, generalizability, and consequential aspects of construct validity. These general validity criteria can be specialized for apt application to performance assessment, if need be, but none should be ignored" (p. 41).

One of the difficulties associated with understanding authentic assessment conceptually stems from the breadth of the assessment approaches that are currently being implemented, as well as the similarity among some of the terms. The issues are both substantive and rhetorical. Substantive issues have to do with such matters as psychometric practice, qualitative/quantitative distinctions, and the relationship between learning and assessment. At the rhetorical level, there is a general lack of precision related to what has become an almost interchangeable use of terms such as authentic, alternative, and performance assessment. Considerable work remains to be done to clarify the conceptual and practical distinctions among these terms (and associated practices).

Another factor that militates against gaining a better understanding of authentic assessment is that not all of the approaches that are being used can be categorized exclusively into discrete categories. For example, portfolios have been promoted as one viable method for making assessment more authentic. But all portfolios are not designed to document authentic learning activities. It is quite possible for

portfolios to contain relatively little that could be classified as authentic. In reality, portfolios typically contain a mixture of authentic and traditional assessment materials. In addition, there are many methods and tools used during the assessment process that cut across assessment categories, such as rubrics, observations, and self- and peer evaluations. But it is important to understand that just because a scoring tool such as a rubric is applied, the assessment is not automatically authentic. The key is to place the emphasis on the authenticity of the activity and whether the assessment strategy appropriately reflects the ability of students to apply what they have learned outside the classroom.

In summary, authentic assessment can involve a mixture of authentic learning and authentic assessment experiences. The first step is to develop activities that require students to apply, integrate, and synthesize knowledge and skill in a manner that reflects the real world and transcends the classroom. Apprenticeships and work study programs are exemplars of approaches vocational educators have used that are set in authentic learning environments. Similarly, it is also expected that assessment strategies should reflect the real world and that they should align with instructional goals and learning experiences. Authentic assessment experiences should, to the extent possible, not be contrived and will often involve multiple measures across time to provide a comprehensive picture of students' knowledge and abilities. It is best to conceive of authenticity as a continuum, representing activities that are totally contrived at one end to those that reflect the real world on the other.

There is currently some disagreement in the literature regarding what sort of standards should be used to gauge authentic assessment. Whereas Hipps (1993) calls for a new set of standards based on constructivist learning theory, Reckase (1997) and Messick (1992, 1996) support the need to retain, and perhaps refine and recast, traditional measurement standards such as validity and reliability. Although this issue is still up for debate, measurement standards, when reported in the research, are predominately discussed in traditional measurement terms.

Authentic Assessment—The Claims

Proponents of authentic assessment have made a variety of claims. Most of these claims fall within two broad categories: improved assessment and improved learning. These are addressed in turn.

Authentic Assessment as a Means of Assessment

It is difficult to discuss alternative assessment without using traditional assessment approaches as a frame of reference. Throughout the authentic assessment literature, there is a rather clear bias against traditional assessment approaches, which typically rely heavily on multiple-choice test items. This perception tends to be reinforced by the fact that nearly every state now mandates standardized testing (Henderson and Karr-Kidwell 1998), which relies heavily on such closed response test items. These tests are influencing educational practices because, in some instances, results are being used as indicators of teacher job performance and are

subsequently affecting teachers' salaries. Critics argue that this practice has resulted in a narrowing of the curricula, due to some teachers' resolve to teach to the test, thus corrupting the entire teaching-learning process (Henderson and Karr-Kidwell 1998; Shepard, Flexer, Hiebert, Marion, Mayfield, and Weston 1994).

Proponents of authentic assessment also worry that traditional forms of assessment (including tests and quizzes) fail to provide a holistic "picture" of student performance and knowledge over time. Traditional measures are designed to yield "snapshots" of what learners know at a given moment. To exacerbate the problem, many of the procedures used to prepare for these types of "snapshot" assessments tend to militate against learning transfer, synthesis, and retention (i.e., cramming and focusing on memorizing facts). These approaches typically do not engage students in authentic tasks and they tend to occur in an artificially contrived environment that does not reflect an activity they are likely to be called upon to do in the real world.

Another argument against traditional assessment practices is that there is an excessive emphasis on paper-and-pencil testing, which encourages the memorization of information. This results in higher test scores shortly following a lesson, while sacrificing long-term retention. Therefore, the goal of authentic assessment should be to provide a comprehensive, holistic, and robust "moving picture" of students' learning experiences by weaving assessment seamlessly into the teaching/learning process.

Most claims of improved assessment can be traced to the premise that if an assessment activity more closely resembles real-world practices, it must be more authentic and thus more valid. Simon and Gregg (1993) claim that "assessment becomes part of the instructional process, and vice versa, as planning evolves based on student progress toward goals, thus increasing the validity of such measures" (p. 4).

The Impacts of Authentic Assessment on Learning

Claims about the positive impacts of authentic assessment on teaching and learning are found throughout the literature. These are so common that it would be impossible to discuss them all in a single chapter. A few of the most common are discussed here.

One of the more general and pervasive premises is that learning experiences that reflect real-world activities are more valid. This validity represents more meaningful educational experiences that are proposed to be the driving force behind improved learning. "The expected positive effects of performance assessments on teaching and learning follow from their substantive validity" (Shepard et al. 1994, p. 6).

Another claim made by both researchers and educators is that authentic assessment experiences can improve student learning (Darling-Hammond, Ancess, and

Falk 1995; Shepard et al. 1994). Many of these claims are closely associated with a constructivist view of knowledge generation. The California Assessment Collaborative (1993) suggests that authentic assessment activities engage students in instructional tasks that require them to construct meaning. Simon and Gregg (1993) indicate that authentic assessments can "stimulate critical thought and input" (p. 6), which suggests that students are engaged in developing higher order thinking. Simon and Gregg (1993) also assert that authentic assessments "involve students in their own learning" (p. 6). These claims parallel those made for cognitive- and metacognitive-based approaches to learning.

Arguments have also been made that authentic assessment experiences encourage multiple modes of expression and support collaboration with others (California Assessment Collaborative 1993; Henderson and Karr-Kidwell 1998; Simon and Gregg 1993). Simon and Gregg also opine that authentic assessment can "increase interest" (p. 6) and "improve attitudes" (p. 6).

In summary, the increasing popularity of authentic assessment tends to parallel the displeasure with education's reliance on traditional measurement practices (e.g., standardized achievement tests). Critics argue that assessment should be more closely linked to real-world expectations and that, by reflecting the real world, resulting assessments become more valid. Therefore, validity appears to be at the heart of these claims. It should be noted that a similar concern has been addressed historically in vocational education, where standardized testing practices have been less prominent and where the boundaries between learning and assessment have been less distinct. In short, one distinct feature of vocational education is that validity concerns have been less problematic than in the more traditional academic content areas.

Authentic Assessment—The Research

In addition to the purported benefits of authentic assessment for the quality of student learning, some claims have also been made about the effect of authentic experiences on student interests and attitudes. Unfortunately, a review of the literature reveals a plethora of anecdotal, rather than empirical evidence. Some authors have acknowledged the rhetorical and advocacy-oriented nature of much of what has been written on the topic and have decried the lack of research. Shepard et al. (1994) state, "to date, little research has been done to evaluate the effect of performance assessments on instructional practices or on student learning" (p. 7). Concern has also been voiced about the quality of the research that *has* been done. This concern is illustrated in a review of portfolio research by Herman and Winters (1994). Although portfolio assessment represents only one aspect of authentic assessment, this review targeting the previous 10 years' literature on portfolios speaks volumes to the issue of quality. Herman and Winters found that, "of 89 articles written on portfolio assessment, only seven report technical data or employ accepted research methods" (p. 48). They also reported that "relatively absent is attention to technical quality, to serious indicators of impact, or to rigorous testing of assumptions" (p. 48).

Gillespie, Ford, Gillespie, and Leavell (1996) also conducted a review of the portfolio assessment literature. In that study, articles spanning the previous 5-year period were reviewed. These manuscripts had been published in the ***Phi Delta Kappan***, ***Educational Leadership***, and six other journals as well as two yearbooks. Although there was no attempt to distinguish between findings based on empirical research versus anecdotal reporting, the information provided was insightful. Gillespie et al. reported that "only five of the articles reviewed mentioned reliability and validity." These results do not suggest that authentic assessment and instructional practices are invalid. Rather, although much of the rhetorical and theoretical support of authentic assessment is compelling, there remains little evidence based on empirical research to support the claims.

The Impact of Authentic Assessment on Learning

Metacognition. Metacognition is the self-management of learning by planning, implementing, and monitoring one's own learning. A metacognitive approach promoted in authentic assessment is to have students participate by using self-assessment strategies throughout the teaching-learning process. Hattie, Biggs, and Purdie (1996) conducted a study to explore this approach. A meta-analysis of 51 studies was used to determine the effect of learning skills interventions to enhance learning. Although their analysis was not limited to studies related specifically to authentic assessment, their findings support the value of metacognition. They recommend that "training for other than mnemonic performance should... promote a high degree of learner activity and metacognitive awareness" (p. 131). This finding supports authentic assessment approaches, which call for students to participate actively in self-assessment, thereby maintaining a sense of where they have been and where they need to go.

In another study focused on metacognition and learning, Moss (1997) found that a group of elementary teachers who were exposed to a "systematic self-reflection" process (in this case, using a rubric) outperformed those who attended the same workshop but did not receive the rubric. The systematic self-reflection group tended to set goals, select interventions to match those goals, and exhibit a deeper level of understanding of the content presented. These findings have further implications for intervention practices, which require students to participate by creating assessment criteria and scoring rubrics. This suggests that allowing vocational students to participate in creating criteria for their own assessments may enhance learning.

Contextual Learning. Teaching in real-world contexts (situated learning) is another important thrust of authentic assessment. The findings of Hattie, Biggs, and Purdie's (1996) meta-analysis of learning skills interventions support the benefits of situated cognition. They recommend that training should "be in context" and "use tasks within the same domain as the target content" (p. 131). Flesher (1993) and Johnson (1987) have conducted studies on the influence of contexts when troubleshooting faults in electricity/electronics. In both studies, the results clearly support the positive influences of context on troubleshooters' abilities to locate faults.

The Value of Authentic Assessment as a Means of Assessment

Student Self-assessment. One of the claims that has been used to support authentic assessment is to reduce the barriers between learning and assessment. One method of doing this is to increase student involvement in their own assessment. Falchikov and Boud (1989), in a meta-analysis of 51 studies related to student self-assessment, explored the relationship between students' self assessment (self-ratings) and their teacher's ratings. It should be noted that the studies included in their review were restricted to those providing quantitative data. The findings indicate a direct relationship between the quality of the design of the study and success of students' self-ratings. Although this illuminates the importance of designing high-quality studies, one could also infer that it is equally important to design high-quality educational activities used for authentic assessments. Another significant finding was related to the experience of the student assessors. Regarding experience and maturity, Falchikov and Boud reported that year in school (i.e., freshman, sophomore, etc.) was not found to be a significant factor in the general quality of students' self-assessments. Self-assessments of students in advanced-level courses more closely resembled their teachers' assessments than those of students in introductory courses. Therefore, when it comes to self-assessment, experience in a given field seems to be more influential than year in school.

Another interesting finding by Falchikov and Boud (1989) was that the category they termed the "broad area of sciences" produced more accurate self-assessments than did the social sciences. Although it is interesting to speculate about reasons for this, it is clear that the types of assessment experiences were relatively similar between the two groups. Also, no patterns existed to signify a difference between assessments of processes versus assessments of products. Neither were there differences between assessment of "professional practices" versus "traditional academic activities." This last finding has direct implications for authentic assessment. "Professional practices" reflect real-world activities called for in authentic assessment. This study suggests that students do no better or worse self-evaluating these activities than they do "traditional academic activities."

Teachers' Level of Performance. Another area of research has focused on how teachers are performing in the classroom. If teachers are not engaging in appropriate forms of authentic assessment, how can students be successful? Haydel, Oescher, and Banbury (1995) conducted a study designed to assess classroom teachers' performance assessments. Ninety-two performance assessments were collected from 79 teachers in a school district that was implementing outcome-based education in Louisiana. Teachers were found to have difficulty following good practices, such as defining purposes and targets and subsequently aligning the two. They also had problems articulating the performance criteria, specifying an appropriate scoring scale, and using a scoring record. It is important to note that this was a single case study, conducted in one school district. Thus, the results may not be generalizable to other populations. However, one could infer that, based on the results of this study along with the findings of Falchikov and Boud's (1989) reported in the previous section, preservice and inservice teacher

training in authentic assessment techniques and practices is likely a key factor in successful implementation.

Reliability and Validity. As noted in the beginning of this chapter, there is some debate in the field as to whether traditional psychometric practices (e.g., those used to establish validity and reliability) are appropriate for authentic assessment. However, given that these practices have a strong history in assessment and psychometrics, related research is examined in this chapter.

One important issue related to assessment is the ability to conclude, with confidence, that what is being reported is consistent and accurate. If policy decisions are to be made based on assessment data, it is important that the reliability and validity of the assessments be established. Gillespie et al. (1996) examined articles on portfolio assessment published over a 5-year span and found that only five mentioned reliability and validity. They thus concluded that the validity and reliability of portfolio assessment (at least for the studies examined) was "controversial at best" (p. 485).

Jiang, Smith, and Nichols (1997) conducted a meta-analysis of 22 studies published after 1980 that were found in the *Educational Resources Information Center* (ERIC) and *Psychology Literature* (PSYCHLIT) databases. The purpose of their work was to identify significant sources of measurement error influencing the reliability of performance assessment. They reported that the number one source of measurement error was due to differences in task difficulty. Further, they found that the complexity of many performance assessments often leads to multiple correct solutions. For example, in a design class, not all design problems are of equal complexity. Even when students are given the same design problem, they often come up with several different plausible solutions. Differences in tasks that have various possible levels of complexity (such as those prevalent in a design class) were found to be the most prominent source of measurement error.

The second most prominent source of measurement error was due to "occasion." Occasion was defined as "all possible occasions on which a decision maker would be equally willing to accept a score on the performance assessment" (ibid., p. 3). If students in a class have the freedom to choose among multiple opportunities when they are to be assessed, there will be greater opportunities for variance in grades due to measurement error. For example, if each student in a vocational welding program is allowed to choose when they are to perform a weld for a grade, there will be a greater chance for variability in grades due to measurement error.

One of the most significant findings reported by Jiang et al. (1997) was that human judgment contributed only a small amount of measurement error. They suggested that it is time that critics set aside their concerns about professional judgments involved in scoring performance assessments. Rather, their findings indicate that error due to human judgment can be minimized through training.

Another study investigated the concurrent validity of performance measures. Crehan (1997) attempted to validate a new performance measure used in a school district by investigating correlations with a norm-referenced achievement measure

previously adopted by the district. He found no significant correlation between the performance measure and the standardized test. Although the appropriateness of using a standardized test as the validity criterion for a performance measure could be questioned, the standardized test was already accepted as a useful predictor of achievement. This is interesting considering that a major reason given for developing performance tests is the claim of the inherent limitations and weaknesses of traditional testing approaches. If one accepts this premise, then the use of standardized tests to validate performance tests could be questioned.

Parkes (1997) conducted a study that addressed the validity of a variety of testing formats with an emphasis on implications for metacognition. He attempted to determine if a student's perceptions of control could be detected during a performance assessment. The hypothesis was that a performance assessment would provide additional information regarding student control whereas a traditional objective test on the same content would not. The findings indicated that the performance assessment score was significantly correlated to the objective test score. This finding supports the contention that they both measured similar content. The findings also indicate that the internal control scale was significantly correlated to the performance score but not significantly correlated to the objective test score. The question posed then was Did the variance due to students' perception of internal control fall within what Messick (1992, 1996) referred to as construct irrelevant variance? Was it extra noise that needs to be controlled for during the assessment process or was it construct relevant variance that is a key part of what was trying to be measured? The researcher concluded that the objective test score measured domain knowledge whereas the performance test better measured ability to use or apply that knowledge. Because of this, Parkes (1997) concluded that "the question now is not which format is more valid, but which construct is the one we really want" (p. 10).

Summary

One of the strengths of authentic assessment is the ability to embed learning within meaningful contexts. Based on this review of the research, this contention can be supported. Teaching in context, a practice that is pervasive in vocational education, can enhance learning. This confirms what career and technology educators have known for years. What is valuable here is to have the importance of authenticity validated in areas that extend beyond vocational and technology-related areas.

The materials reviewed in this chapter also support the value of metacognitive approaches to learning and assessment. Encouraging students to become more involved in monitoring their own learning through self-evaluation can enhance student learning. From the assessment side, research indicates that students do a better job evaluating their own work in upper-level classes in a given field than entry-level classes. This could be due to a number of factors, such as maturity or additional content knowledge. Additional research is needed to explore the use of self-assessment in vocational subjects.

One significant concern throughout this review had to do with teachers' performance, both as facilitators of learning and as evaluators of student performance. Research indicates that teachers may have difficulty maintaining alignment among performance criteria, scoring scales, and the assessment records. This finding supports the need for better preservice and inservice training in authentic assessment and contextualized learning practices. It is important to note that the in-service needs of vocational teachers will likely be quite different from those of academic teachers. Contextualized learning and many authentic assessment practices are not new to vocational teachers. There is, however, an ongoing need for vocational teachers to understand how authentic assessment mechanisms work as well as how to integrate learning with academic areas.

One of the key measurement issues discussed throughout the research had to do with reliability. The two largest sources of measurement error in performance assessments were differences in difficulty of tasks and variance due to multiple occasions in which teachers are equally willing to accept a score. These represent relatively straightforward psychometric issues that must be addressed in any type of research. However, both concerns tend to be exacerbated when the emphasis shifts away from testing to context-based, authentic assessment techniques. One of the surprising findings in this review was that human judgment emerged as a less serious, and correctable, source of reliability error than might have been expected. The research indicates that proper training can minimize human judgment error. This indicates that the scoring and use of authentic assessment measures are appropriate topics for teacher inservice training.

Finally, the question of the nature of authenticity was addressed. How authentic is authentic enough? Although research indicates that context can have a positive influence on learning, there was a general lack of research investigating the ranges of authenticity. How closely does education need to mirror the real world in order to have positive impacts on learning and assessment? Is there a point of diminishing returns? Is it possible for an activity to reach a threshold of authenticity beyond which it is no longer prudent to expend the resources required to increase its effectiveness? Do all of our educational activities have to reflect the "real world"? Are there some aspects of the curriculum where learning occurs better using traditional approaches? Research remains to be done in these areas.

Authentic assessment represents an exciting attempt to stimulate learning and make it more relevant. It also represents a means for assessing students in rich and meaningful ways. Students deserve to know why it is important to learn something, and authentic teaching and evaluation methods represent a move in that direction. However, it is important to note that authentic assessments should represent only one category of tools and, like all tools, should probably not be used exclusively for all tasks.

The purpose of this chapter was to attempt to identify what the research has to say about authentic assessment. As is frequently the case in education, the linkage between practice and research is often tenuous. Trends tend to come and go. The current enthusiasm and interest in cognitive learning theory, with its emphasis on authentic learning and assessment, represents a special opportunity for vocational education. Other academic areas are coming to realize what vocational educators have known to be true for years: meaningful, contextualized experiences tend to promote better learning. The challenge remains to engage and focus the best minds in the profession to conduct the research needed to clarify how these mechanisms work...and don't work.

Authentic Assessment Tools

John Scott
The University of Georgia

Skillful and effective teachers require students to analyze and synthesize information, apply what they have learned, and demonstrate their understanding of material according to specified criteria. They have developed learning and assessment experiences to engage students and teach them how to "produce," rather than simply "reproduce" knowledge (Burke 1992, p. 5). In these classrooms, the emphasis shifts from facts and isolated knowledge to active learning, where students work together to examine information and issues, solve problems, and communicate ideas. These shifts in emphasis are often accompanied by changes in assessment practices typified by involving students in authentic tasks, measuring a variety of outcomes, and involving students in self-assessment and reflection.

The focus of this chapter is on the "tools" used to conduct authentic assessment. It is important to preface this discussion by thinking about some key contextual issues. As anyone who has ever worked with tools of any kind knows, tools can be (and often are) misused. They are often used in ways and for purposes other than those for which they were designed. To press the analogy still further, most "tool boxes" contain a diverse selection of tools, each of which are selected and used for various purposes. Appropriate tool selection and use is a function of the knowledge and skill of the "tool user." Much the same is true of authentic assessment. The toolbox is full of tools; but we must first think carefully about the various contexts and purposes for which they are used.

Connecting, Reflecting, and Feedback

There are three important aspects or concepts that should accompany any type of authentic assessment: connecting, reflecting, and feedback.

Connecting

Across the nation, considerable attention is being directed toward the reform of testing and assessment. Much of this thrust is designed to extend assessment beyond testing, with its emphasis on facts and fragments of information, to authentic methods of assessment. A key feature of many of these authentic strategies is that students are required to connect facts, concepts, and principles together in unique ways to solve problems or produce products. Cognitive research has challenged the belief that learning and learning transfer occur simply by accumulating and storing bits of information (Shepard 1989, p. 4). Contemporary learning theory holds that learners gain understanding as they draw on and extend previously learned knowledge, construct new knowledge, and develop their own cognitive maps (connecting diagrams) interconnecting facts,

concepts, and principles. Research indicates that information learned and assessed as a linear set of facts fails to yield the kinds of in-depth understanding needed to function in our modern society.

Glaser (1988) describes a number of different types of evidence collected through assessment. One of the most important of these is "coherence of knowledge." Glaser goes on to observe that beginners' knowledge is spotty and superficial, but as learning progresses, understanding becomes integrated and structured. Thus assessment should tap the connectedness of concepts and the student's ability to access interrelated chunks.

Authentic assessments are almost always framed in the form of learning experiences. These experiences are typically sequenced from simple to complex and are progressive in nature. An important role of teacher-facilitators is to help students connect the knowledge and skills learned in previous tasks and then extend them to related or more complex tasks. Transfer of knowledge and skills is enhanced when students recognize the connectedness of learning. A number of authentic assessments such as graphic organizers, writing samples, and portfolios require students to connect (or synthesize) what they have learned to produce finished products. Many technical tasks presented in technology-based programs require students to connect their previous knowledge of mathematics, science, social studies, and English to solve problems and complete tasks and projects.

Reflecting

The range of available options for teachers wishing to improve student assessment extends beyond the cognitive and psychomotor domains to include assessment of attitudes and other affective behaviors. The key element here is to help students develop their self-awareness and reflective skills. Students need to learn how to assess their own work and to think about their thinking. A key aspect of many forms of authentic assessment is the opportunities that are provided for students to reflect on their thinking, practices, and learning. The technical term for this type of reflective process is metacognition.

Robin Fogarty (1994), in her excellent book ***The Mindful School: How to Teach for Metacognitive Reflection***, defines metacognition as a sense of awareness—"knowing what you know and what you don't know" (p. viii). Barell (1992) extends Fogarty's definition to include feelings, attitudes, and dispositions because thinking involves not only cognitive operations but also the dispositions to engage in cognitive activities.

Burke (1994) notes that metacognitive reflections provide students with opportunities to manage and assess their own thinking strategies. "Metacognition involves the monitoring and control of attitudes, such as students' beliefs about themselves, the value of persistence, the nature of work, and their personal responsibilities in accomplishing a goal" (p. 96). These attitudes are fundamental to all tasks in varying degrees, whether academic or nonacademic. Teachers need to provide opportunities for students to engage in the kind of metacognitive monitor-

ing where they reflect on "what we did well, what we would do differently next time, and whether or not we need help" (p. 96).

Numerous researchers (Barell 1992 Fogarty, Perkins, and Barell 1992; and Perkins and Salomon 1992) have explored the critical relationship between metacognition and learning transfer. Barell (1992) states that "in order to transfer knowledge of skills from one situation to another, we must be aware of them; metacognitive strategies are designed to help students become more aware" (p. 259). Fogarty, Perkins, and Barell (1992) define transfer as "learning something in one context and applying it in another" (p. ix).

In the constructivist view of learning, individuals absorb information and make sense of that information through metacognitive reflection. Reflection allows individuals to recognize the gaps that exist in their understanding. As gaps are recognized and become significant to students, they are motivated to locate, apply, and connect previous learning as well as to construct new knowledge.

Burke (1994) and Fogarty (1994), in their works on metacognition, detail a number of metacognitive strategies that can be used by classroom teachers. These include such techniques as Mrs. Potter's Questions, KWL charts, PMI charts, transfer journals, wrap-around, reflection page, learning logs, seesaw thinking, pie in the face, stem sentences and many others.

- Mrs. Potter's questions: What were you expected to do in this assignment? What did you do well? If you had to do this task over, what would you do differently? What help do you need from me?

- The KWL strategy consists of a three-column chart in which one column (K) is devoted to what I **Know**, the second (W) to what I **Want** to know, and the third (L) to what I **Learned** after finishing this lesson or assignment.

- The PMI strategy is similar to the KWL chart except the first column (P) is devoted to the **Plus** or favorable things found about a learning experience, the second (M) focuses on the **Minuses** or unfavorable finding, and the third (I) is devoted to what the student found **Interesting** about the learning experience.

Descriptions of other metacognitive strategies can be found in Burke's and Fogarty's books. It is very important to provide opportunity for learners to reflect on what has been learned as teachers rush to "cover the content in the textbook" and prepare learners to "pass the test." Many learners are unaware of their thinking processes while they are learning and trying to create personal meaning out of some learning experience. When asked to describe what they initially thought about a topic, how they began to create personal understanding about some content, and what they would be able to do with this new knowledge or skill, they can't describe how they went about it and usually reply "I don't' know how I did it, I just did." Students who are taught how to reflect on learning by using metacognitive reflection strategies should be able to monitor, assess, and improve their own thinking and learning performance.

Feedback

Another important outcome of authentic assessment has to do with providing feedback to learners related to significant objectives. Wiggins (1993) notes that many teachers erroneously believe they are providing feedback with test scores and coded comments such as "good work," "vague," and "awkward." What is wanted and needed by learners is user-friendly information about performance and how improvement can be made. Learners need information that will help them self-assess and self-correct so that assessment becomes integrated throughout the learning experience.

Wiggins (1993) draws a subtle, but important, distinction between guidance and feedback. Guidance gives direction whereas feedback tells one whether or not they are on course. Guidance is typically teacher initiated and tends to be prescriptive. By contrast, feedback actively involves and engages the learner. Frequently, the process is collaborative and reflective; the teacher and student become partners in the learning process. Figuratively, feedback techniques are those experiences that help students see themselves and their performance more clearly. Throughout the assessment process, students are provided with real-time information about the quality of their performance.

Wiggins (1993) notes that feedback is more like a running commentary rather than measurement. It enables learners to monitor their performance, thinking about whether or not they are on the right track without labeling or censoring their performance. From this feedback perspective, the emphasis shifts from "measurement" as an end goal to "assessment" as an ongoing and continuous process. To maximize the effect, feedback should occur while the performance is underway, not just after it is evaluated.

Mastery of complex, integrative learning activities extends well beyond simply responding to probing questions following performance. Rather, it involves continuous feedback throughout the process of solving complex problems. Successful performance requires concurrent feedback inherent in the task itself or in the context in which the task is performed that enables learners to self-assess and self-correct as accurately as possible. Optimally, feedback is best when it becomes an integral part of students' own mental processes, when they learn how to assess themselves. Similar to other real-life situations, feedback is comprised of a complex set of external (family members, friends, co-workers, and supervisors) and internal messages (reflective and metacognitive thinking).

Self-Assessment

One of the more exciting, but underused, dimensions of authentic assessment is student self-assessment. Students want to know how they are doing *while* they are performing some tasks and, even more, they want to know how well they did when the task is completed. In traditional assessment, students must wait until post-performance tests have been graded for feedback. In alternative assessment

classrooms, students are encouraged to engage in self-assessment and to collaborate with teachers to review performance and decide the next steps in the learning process.

One of the key aspects of student self-assessment has to do with criteria (or standards). These criteria come in different forms. In "self-referenced" assessment, learners evaluate performance in light of their own goals, desires, and previous attainments and thus become more cognizant of present performance as well as steps that must be taken to extend their learning. In this type of self-assessment, standards are embedded in the value system and inherent goals of students. In "standards-referenced" self-assessment, learners compare their own characteristics of performance against established standards or criteria.

Self-assessment abilities represent a critical workplace skill. In the workplace, individuals are continuously faced with situations in which they must assess situations, make decisions, and then evaluate the quality of those decisions. This type of authentic, formal self-assessment activity is rare in most public schools and universities. In most schools, students rarely have the opportunity to evaluate their own performance, because teachers have assumed the assessment role. Teachers who bemoan student apathy, lack of personal investment in their own education, willingness to settle for minimal performance, and even cheating may not realize that they are experiencing the results of teacher-vested assessment. What if students could be genuinely empowered to engage in meaningful self-assessment? What if the locus of authority in the assessment process were to be shifted from teacher to student, where the authority is shared? What if students had a real voice in developing and assessing their own learning?

At this point, it is important to acknowledge that this vision of self-assessment is contingent on such things as students' developmental level, maturity, and previous educational experiences. Self-assessment techniques are not uniformly appropriate and will not always work. However, students who are given the opportunity to become more engaged in the learning process and in assessing their own progress often do respond with intelligence, responsibility, and determination after a learning period in which they develop assessment skills (Mabry 1999). For example, D'Urso (1996) reports the results of a study of second-grade students involved in their own assessment. She concludes that students' sense of self improved, their work became more meaningful to them, they became protective of the knowledge they had gained, and they began to reflect on what they knew as well as on what they still needed to discover. They discovered their own "voice" and developed a deeper sense of self.

Strategies and Tools

We now turn our attention to the tools themselves. These tools must be carefully selected to provide opportunities for students to practice and perform meaningful tasks that are reflective of life outside of the classroom. Authentic assessment starts with the selection of meaningful learning tasks. These tasks need to be organized and structured so that they are contextualized, integrative,

metacognitive (require students to think about thinking), related to the curriculum taught, flexible (require multiple applications of knowledge and skills), open to self-assessment and peer assessment, contain specified standards and criteria, and are ongoing and formative (Weber 1999).

Mabry (1999) notes that we must match purpose or outcome expectations with assessment strategies. "What do we want to assess–and do we really need to assess it?" "Why do we want to assess it–what will we do with the results?" "How should we assess–how can we get the information we need?" "How can we assess without harmful side effects?" (p. 41). The central issue here has to do with "tool selection." Given a particular problem, situation, or set of questions, teachers need to learn to ask, "What is the best tool for the job?"

Teachers will need to use a variety of assessment tools and techniques in order to enable all students to have a more complete picture of their growth and achievement. The National Center for Research in Vocational Education study *Using Alternative Assessment in Vocational Education* (Stecher et al. 1997) identified four categories of alternative assessment that are widely used in vocational education: (1) written assessments, including selected response types such as multiple choice and constructed responses types such as essay items or writing samples; (2) performance tasks; (3) senior projects including research papers, performance projects, and oral presentations; and (4) portfolios. With the development of computer-based simulation software, additional possibilities are being developed.

A wide variety of assessment tools are available to teachers and students. As one reviews the list of tools, it will become immediately obvious that there is scant distinction to be made between performance activities and assessment techniques. A key feature of authentic assessment is a "blurring" of the distinctions typically drawn between classroom activities and assessment (see Figure 1).

The kinds of performance activities shown in Figure 1 can serve as a basis for developing authentic assessments to transform assessment practices from summative and teacher directed to formative and student centered. A detailed discussion of each of these performance activities and how to structure assessment components is beyond the scope of this work. However, it is useful to make some general observations about the usefulness of these techniques as well as ideas for implementation. Following the general overviews, three performance activities (learning logs and journals, portfolios, and projects) are discussed in more detail. There is a growing body of well-illustrated resources available that are designed to help teachers structure authentic assessments. One particularly useful resource for authentic assessment tools is Skylight Professional Development <www.skylightedu.com>.

Graphic Organizers and Concept Mapping

- Concept maps
- Data tables
- Cause and effect diagrams
- Graphs
- Run control charts
- Flowcharts
- Pareto diagrams
- Correlation/scatter diagrams
- Idea webs/graphic organizers
- Geographic maps
- Time lines
- Venn diagrams
- Event chains
- Histograms
- PMI strategy reports
- Mrs. Potter's questions
- Connecting elephants
- Big idea generation
- Ranking ladders
- Mind maps

Performance Products

- Business letters
- Autobiographies
- Editorials
- Displays
- Drawings/illustrations
- Experiments
- Essays
- Surveys
- Storyboard reports
- Job applications
- Book reviews
- Bulletins
- Critiques
- Crossword puzzles
- Designs
- Requisitions
- Vitas/Resumes
- Inventions
- Lab reports
- Information-seeking letters
- Management plans
- Math problems
- Geometry problems
- Models
- Writing samples
- Job searches
- Cartoons or comics
- Collages
- Consumer reports
- Handbooks
- Booklets
- Home projects
- Pamphlets
- Observation reports
- Research reports
- Posters
- Workplace scrapbooks
- Grant applications
- Team reports
- Career plans
- Video yearbooks
- Training plans
- Exhibits
- Ballads
- Announcements
- Biographies
- Questionnaires
- Technical repairs

Live Performances and Presentations

- Interviews
- Issues/controversy
- Workplace skits
- Slide shows/video
- Human graphs
- Announcements
- Games/quiz bowls
- Student-led conferences
- Story time/anecdotes
- Prepared and extemporaneous speeches
- Commercials
- Demonstrations
- Newscasts
- Plays-TV/radio broadcasts

Figure 1. *Authentic assessment tools/performance activities*

Graphic Organizers and Concept Mapping

Graphic organizers are visual representations of mental maps using important skills such as sequencing, comparing, contrasting, and classifying. They involve students in active thinking about relationships and associations and help students make their thinking visible. Many students have trouble connecting or relating new information to prior knowledge because they cannot remember things. Graphic organizers help them remember because they make abstract ideas more visible and concrete. This is particularly true for visual learners who need graphic organizers to help them organize information and remember key concepts (Burke 1994).

Teachers can help students use graphic organizers by modeling and using topics that can be easily understood. Students can develop skills in developing graphic organizers if they are allowed to work first in small groups and can select a topic of their choice related to the lesson content.

Although graphic organizers are learning tools, they can also effectively be used as authentic assessment tools. Teachers who involve students with graphic organizers need to develop exemplary models that can be used for assessment. Criteria describing what content and relationships should be visually shown in student work need to be developed and used in rubric (scoring) form to make assessments more objective. Similar to essay questions, which require written expression in a connected manner, graphic organizers require students to present information in written and visual format. Graphic organizers also can be used as a test item format to assess student learning. This provides students with a creative and engaging way of expressing what they know and are able to do.

Performance Products

Many of the performance activities are end products of learning that can be assessed by rubrics (scoring forms) and other assessment tools designed to measure both processes and product quality.

Teachers who use authentic performance products provide students with opportunities to construct knowledge in real-world contexts so they can understand what they have learned. These products serve as a culminating experience in which students can retrieve previous learning, organize important information, and complete an assigned activity showing mastery of what they have learned.

Some teachers are reluctant to assign performance products because they do not feel comfortable grading them. They recognize that it takes time to construct exemplary models and to develop criteria and performance indicators required for rubric development. The key to assessing performance products is to set the standards and criteria in advance. Students who know the criteria that will be used to assess their work receive valuable instructional guidance in completing their products so they meet and/or exceed expectations.

As teachers recognize the importance of engaging students in making performance products, they will learn how to structure the learning environment to facilitate the process. They will also plan ahead to develop the tools needed to assess both the process of developing the product as well as the completed product. Scoring rubrics are one of the key assessment tools used for performance products. Information on how to construct and use them follows later.

Live Performances and Presentations

As with performance products, the key to effective assessment of live performances and presentations is establishing the criteria and performance indicators in advance. Criteria and performance indicators effectively organized into scoring

rubrics provide examples of what students must do to demonstrate that they have learned at a specified level. The most important assessment strategy with live performances and presentations is to engage students in assessing their own performance first, followed by teacher assessment and an opportunity for students and teachers to interact over assessment findings. Live presentations involve two major assessment factors. One is the quality of the assigned work and the second is the demonstration of presentation skills. Scoring rubrics must include both of these factors.

Rubrics

Among the most common methods for student self-assessment are scoring rubrics. Marzano, Pickering, and McTighe (1993) have defined rubrics as "a fixed scale and list of characteristics describing performance for each of the points on the scale" (p. 10). Rubrics are scoring devices (or tools) that are designed to clarify, communicate, and assess performance. They are grading tools containing specific information about what is expected of students based on criteria that are often complex and subjective.

Rubrics typically contain two important features; they identify and clarify specific performance expectations and criteria, and they specify the various levels of student performance. In their simplest form, rubrics are checklists requiring a "yes" or "no" response. More complex rubrics include written standards of expected student performance with different levels of performance indicators describing student performance that meets or exceeds the standard.

There are as many different types of rubrics as there are rubric designers. Most rubrics fall under the two categories, holistic or analytical. Holistic rubrics consider performance as a totality, with the primary purpose being to obtain a global view of performance, typically on complex tasks or major projects. By contrast, analytical rubrics are designed to focus on more specific aspects of performance. Their purpose is to provide specific feedback on the level of performance on each major part, with the advantage of providing a detailed analysis of behavior or performance. These rubrics detect strengths and weaknesses and identify areas for refinement.

Rubrics of both types can be used appropriately for product and process assessment as well as for formative and summative assessment. It is also important to note that rubrics are typically developed and used as open communication devices. For example, it is not unusual for students to be involved in the process of developing the rubrics that will be used to assess their performance. Used in this way, rubrics become an effective mechanism for clarifying and openly communicating the expectations of learning activities. Many teachers share and discuss the contents of rubrics that will be used to assess an activity early in the process. As a result, the expectations are clarified and, in some cases, negotiated.

There are numerous advantages to using rubrics provide for both students and teachers:

- Enabling assessment to be more objective and consistent,
- Focusing attention of the assessor on the important outcomes with an assigned value for each,
- Demystifying the expectations for the student by assigning values for each expected outcome,
- Allowing students to identify strengths and to focus on weak areas while providing opportunity to revisit them,
- Prompting teachers to identify critical behaviors required for task completion and to establish the criteria for performance in specific terms,
- Encouraging students to develop a consciousness about the criteria they are to demonstrate in their performance as well as the criteria they can use to assess their own abilities and performance,
- Promoting an emphasis on formative as well as summative evaluation,
- Providing benchmarks against which to measure and document progress,
- Lowering student anxiety about what is expected of them,
- Ensuring that students' work is judged by the same standard, and
- Leading students toward high-quality performance.

There are some disadvantages as well. Rubrics can be time consuming to develop and use. Good rubrics also must be grounded in clearly identified and stated criteria or standards. In many cases, these have not yet been identified or developed. Once the criteria have been clarified, considerable work remains to clearly identify the key indicators that will be used to assess the various levels of attainment for each of the criteria. This is the hard work of solid, clear, and meaningful assessment. The expectations must be clarified and then the level of attainment must be described and clearly communicated.

Some general guidelines for involving students in constructing and using rubrics have been developed by Goodrich (1997):

1. Begin by looking at models. Show students examples of good and not so good work. Identify the characteristics that make the models good and the bad ones bad.

2. List the critical criteria for the performance. A good guide is to think about what you would need to include if you had to give feedback to a student who did poorly on a task. Students can be involved in discussing the models to begin a listing of what counts in high-quality work.

3. Articulate gradations of quality or determine the quality continuum. Describe the best and worst levels of quality, and then fill in the middle based on knowl edge of common problems associated with the performance. Use descriptive terms such as Not yet, OK, and Awesome instead of failure, average, and excellent.

4. Engage students in using the rubrics created to evaluate the models given them in step 1 as practice in self-assessment and to pilot test the rubrics.

5. Give students their task. As they work, stop them occasionally for self- and peer assessment using the rubrics provided.

6. Give students time to revise their work based on the feedback they received in step 5.

7. Use the same rubric students used to assess their work. This is made possible by including a scoring column for students, peers, and teachers.

8. Schedule a debriefing time with students to compare their rubric scoring with those completed by the teacher. Require students to reflect on the next steps in the learning process.

One excellent resource is ***Assessing Student Outcomes: Performance Assessment Using the Dimensions of Learning Model*** by Marzano, Pickering, and McTighe (1993), published by the Association for Supervision and Curriculum Development. This work contains many examples of rubrics for specific tasks and situations. Another approach to developing rubrics using a "shell" to cluster criteria according to valued workplace competencies (e.g., creative thinking, contributing citizen, problem solving, effective communication, etc.) was developed by Custer (1996).

Portfolios

Another alternative assessment tool that has attracted widespread popular attention is portfolios. Portfolios are collections of student work gathered over time. The contents of portfolios can range from comprehensive coverage containing a plethora of materials to those that are quite selective, containing only a limited number of student-selected items. Student portfolios offer a range of flexibility that makes the method attractive to a wide range of teachers and programs. The elements to be included in this type of assessment are almost endless. Several critical components of effective portfolios are—

- A thoughtful student-developed introduction to the portfolio,
- Reflection papers behind each major assignment of the portfolio,
- Scoring rubrics for portfolio entries that enable students to self-assess their work,
- Established models, standards, and criteria that enable students to select their best work to be included in the portfolio, and
- Student oral presentation of their portfolios to significant others such as peers, teachers, and parents.

Portfolio assessment offers many advantages, but Frazier and Paulson (1992) note that the primary value of portfolios is that they allow student the opportunity to evaluate their own work. Further, portfolio assessment offers students a way to take charge of their learning; it also encourages ownership, pride, and high self-esteem. Portfolios can be maintained over several years and can be used as "pass-

ports" as students move from one level of education to another. Portfolio passports can also be used as valuable tools for obtaining jobs in business and industry.

Portfolio assessment requires careful thought and preparation on the part of both teachers and students. Vavrus (1990) offers the following considerations and recommendations that should be considered in designing a portfolio assessment system.

- ***What will it look like?*** Portfolios must have both a physical structure (binder as well as the arrangement of documents within the portfolio) and a conceptual structure (underlying goals for student learning).

- ***What goes in?*** To answer this question, other questions must first be addressed: Who is the intended audience for the portfolios? What will this audience want to know about student learning? How will these audiences be involved in portfolio development? Will selected documents of the portfolio show aspects of student learning that traditional test results do not show? What kinds of evidence will best show student progress toward expected learning outcomes? Will the portfolio contain best works only, a progressive record of student growth, or both? Will the portfolio include more than finished pieces–for example, notes, ideas, sketches, drafts, and revisions?

- ***How will procedural and logistical issues be addressed?*** How will student working files and portfolios be kept secure? When will students select documents to include in their portfolios? When will some portfolio document be taken out to specialize the portfolio? What criteria or assistance will be provided to students so that they can reflect on their work, monitor their own progress, and select pieces for inclusion in the portfolio? Will students be required to provide a rationale or explanation for work selected for inclusion in the portfolio?

- ***How will portfolios be evaluated and who will be involved?*** It is critical that students be actively involved in assessing their own work. To facilitate student self-assessment teachers will have to answer some important questions. What factors will be evaluated such as achievement in relation to standards, student growth along a continuum, or both? What models, standards, criteria and instruments will have to be developed to guide assessment? When will portfolio entries be evaluated? Will other teachers be involved assessing portfolio elements? Will parents or guardians be involved in assessing the portfolio? If so, how?

- ***What will happen to the portfolio at the end of the semester or school year?*** Will they be turned over to students at the end of the course or school year to keep and use as they see fit? Will students be encouraged to keep their portfolios over an extended period of time and use them as "passports" for entry into other levels of education or to work?

It is clear that portfolios are a way of collecting and packaging a comprehensive body of rich evaluation materials. The key is to think carefully through the many logistical, conceptual, and procedural issues that must be addressed in order for this tool to be used effectively. Portfolios should not be "a place to dump anything and everything" loosely related to a given course. Rather, their value as an assessment tool is maximized when they contain items that have been carefully and thoughtfully selected to address specified learning goals. At their best, portfolios can represent an extremely rich portrait of student ability and interest.

Learning Logs and Journals

Learning logs and journals are tools designed to cause students to reflect on what they have learned or are learning. Used properly, they encourage student self-assessment and provide a mechanism for making connections across the various subject matter areas. Journals have been used widely in English classes for many years. Now they are being adopted by other teachers to develop communication skills and to help students to make connections, examine complex ideas, and think about ways to apply what they have learned over an extended period of time. Herman, Aschbacher, and Winters (1992) indicated that the fundamental purpose of learning logs and journals is to "allow students to communicate directly with the teacher regarding individual progress, particular concerns, and reflections on the learning process" (p. 2).

A distinction can be made between learning logs and journals. Learning logs usually consist of short, objective entries under specific heading such as problem solving, observations, questions about content, lists of outside readings, homework assignments, or other categories designed to facilitate recordkeeping (Burke 1994). Student responses are typically brief, factual, and impersonal. Fogarty and Bellanca (1987) recommend teachers provide lead-ins or stem statements that encourage students responses that are analytical (breaking something down into its parts), synthetic (putting something together into a whole), and evaluative (forming judgment about the worth of something). Example log stems include the following: One thing I learned yesterday was..., One question I still have is..., One thing I found interesting was..., One application for this is..., and I need help with...

By contrast, journals typically include more extensive information and are usually written in narrative form. They are more subjective and focus more on feelings, reflections, opinions, and personal experiences. Journal entries are more descriptive, more spontaneous, and longer than logs. They are often used to respond to situations, describe events, reflect on personal experiences and feelings, connect what is being learned with past learning, and predict how what is being learned can be used in real life (Burke 1994). As with learning logs, stem statements can be used to help students target responses. Example lead-ins are as follows: My way of thinking about this is..., My initial observation is..., Upon reflection I...

Learning logs and journals can be used in the following ways (Burke 1994):

- Record key ideas from a lecture, video, presentation, field trip, or reading assignment,
- Make predictions about what will happen next in a story, video, experiment, event, situation, process, or lesson,
- Record questions and reflect on the information presented,
- Summarize main ideas of a lesson, article, paper, video, or speech,
- Connect the ideas presented to previous learning, or to other subjects or events in a person's life,
- Monitor change in an experiment or event over time,
- Brainstorm ideas about potential projects, papers, presentation, assignments, and problems,
- Help identify problems and record problem-solving techniques, or
- Track progress in solving problems, readings, homework assignments, projects, and experiences.

Learning logs and journals can be effective instructional tools to help students sharpen their thinking and communication skills. They give students the opportunity to interact with the teacher, lesson content, textbooks, and each other. They also afford students an opportunity to think about material, clarify confusion, discuss key ideas with others, connect with previous learning and experiences, and reflect on the personal meaning of subject matter. They provide a record over time of what has been presented and learned. Furthermore, logs and journals are typically best used to promote formative assessment, although they also can be structured to provide summative assessment information.

Projects

Many different types of projects can be developed to challenge students to *produce* something rather than *reproduce* knowledge on traditional tests. Projects allow students to demonstrate a variety of skills including communication, technical, interpersonal, organizational, problem-solving, and decision making skills (Burke 1994). Projects also provide students with opportunities to establish criteria for determining the quality of the planning and design processes, the construction process, and the quality of the completed project.

The Southern Regional Educational Board has published a guide to preparing a syllabus for its *High Schools that Work Program* that includes a major focus on projects as the centerpiece of curriculum, instruction, and evaluation. This guide, ***Designing Challenging Vocational Courses*** by Bottoms, Pucel, and Phillips (1997), describes the procedures required to select and sequence major course projects, develop project outlines, decide on an instructional delivery plan, and develop an assessment plan.

Several states, notably California and Kentucky, have made successful completion of a student-initiated culminating project (senior project) a part of their student assessment system. The California Department of Education (1994), in collabora-

tion with the Far West Laboratory, has developed the Career-Technical Assessment Program (C-TAP), which includes a C-TAP project. The project is a major piece of "hands-on" work designed and completed by each student. The project becomes an instructional and assessment tool that allows students to demonstrate skills and knowledge learned in a sequenced instructional program. Completing the project provides a mechanism for students to plan, organize, and create a product or event. Through this process, students are able to pursue their own interests, meet professionals in the field who can offer advice and instruction related to their project, work cooperatively with others in certain parts of the project, and apply the knowledge and skills they have learned in other school subjects. Each student's project must be related to the career-technical program in which they are enrolled and can take as little as a few weeks to complete or several months. Students are allowed to work on the project themselves or in small groups. There are four major sections of the C-TAP project:

1. *Plan:* A process that helps the student design the project
2. *Evidence of Progress:* Three pieces that show the student's progress toward developing the final product
3. *Final product:* A final product that is the result of the student's work
4. *Oral presentation:* An oral presentation in which the student describes the project, explains what skills were applied, and evaluates his or her work

C-TAP projects are evaluated in two ways with two separate scores being generated. First, the project is rated using a rubric focused on three evaluation dimensions: content, communication, and responsibility. Content pertains to career-technical knowledge and skills, communication relates to the overall presentation of work, and responsibility pertains to the student's ability to complete work independently. The second score (also generated using a rubric) focuses on oral presentation skills including public speaking skills, content knowledge, and analysis. A student manual and a teacher guidebook contains the information necessary for the complete operation of the C-TAP program.

Summary

Many factors are driving assessment reform in this country, including an emphasis on constructivism and authenticity, standards, and higher-order thinking skills. These forces and others have stirred interest in the educational community to look for alternatives to traditional testing in order to give a more accurate and complete picture of student growth and achievement. Organizations that specialize in assessment (e.g., the Far West Laboratory and the Center for Research on Evaluation, Standards, and Student Testing) are working with school systems to develop and test alternative assessments. The preliminary results are quite promising in terms of reform in curriculum and instructional practice as well as increased student engagement in the learning and assessment process. Assessment of learning is truly a "work in process." It is exciting to see the progress that has been made to move beyond teaching and testing fragmented lists of declarative knowledge in favor of involving students in applying knowledge in unique and authentic ways.

The challenge for teachers is to commit to change the way they teach and assess students as well as put forth the effort to develop and use alternative assessment strategies such as those described in this chapter. Every effort should be made to develop meaningful, authentic learning and assessment tasks that target the knowledge, skills, and attitudes necessary for learning and life. Educators must also learn how to organize and structure these tasks so that they are contextualized, integrative, flexible, and open to self-assessment and peer assessment. Additionally, a clear focus on standards and criteria must be maintained in a way that provides for both formative and summative procedures. Students should be encouraged to become actively involved in the assessment process through metacognitive reflection, establishing criteria and performance indicators required to develop effective scoring rubrics, and using these scoring instruments to assess their own work. Effective feedback is the key to improved student learning. Yet many teachers are reluctant to spend the time required to develop and exhibit exemplary models of expected performances and to teach students how to assess and regulate their own performance.

Considerable progress has been made in the 1990s in designing and implementing alternative assessments. There are many success stories that point toward systemic change in the way educators are structuring curriculum, delivering instruction, and assessing student growth and achievement. Much of this work closely mirrors work that has been done in vocational education for many years. The current shared interest between the vocational and academic communities holds promise for improving both as teachers share ideas, techniques, and tools across disciplines.

Authentic assessment supports change in curricula, teaching, and school organization. But the real question is "Do these new assessment methods and techniques contribute to improved student learning?" A growing number of teachers seem to think so. Reporting on the effects of authentic assessment in action at five schools, Darling-Hammond, Ancess, and Falk (1995) note that classroom interactions, student work, exhibitions, and hallway conversations provide widespread evidence of in-depth learning, intellectual habits of mind, high-quality products, and student responsiveness to rigorous standards.

Large-Scale Authentic Assessment

Marie Hoepfl
Appalachian State University

Student assessment is the most widely used approach being taken by state and federal policymakers in their attempts to leverage improvement of instruction in the nation's schools. In particular, performance assessment is seen as a way to set targets for what students should know and be able to do, encourage curricular reform, and improve teaching methods. Among other claims, performance assessments are viewed as mechanisms for promoting greater educational equity (Roeber 1997). The reliance on standardized, norm-referenced multiple-choice tests for large-scale assessments has yielded recently to an increased emphasis on performance skills and "thinking abilities" needed in the workplace and in daily life. Proponents claim that performance assessments can better tap the skills and abilities that students need (Darling-Hammond, Ancess, and Falk 1995).

Since 1970, when standardized tests began to be more widely used, educational researchers have seen slight increases in basic skills test scores, but declines in measures of higher-order thinking skills. Officials within national organizations, ranging from the National Research Council to the National Councils of Teachers of English and Mathematics, among others, have attributed this decline to the emphasis on tests of basic skills, which have driven the curriculum (ibid.).

The structure behind performance assessment contrasts sharply with the discrete items found on multiple-choice assessments. Rather than artificially separating desired knowledge and skills into small pieces, performance assessment attempts to measure behavior as an intact whole (Yen 1993). "In assessment reform theory, all performance assessments must require students to *structure* the assessment task, *apply* information, and *construct* responses, and, in many cases, students must also be able to *explain* the processes by which they arrive at the answers" (Khattri, Reeve, and Kane 1998, p. 2).

The latest rounds of curriculum reform advocate the use of performance assessment as a lever for encouraging curricular and instructional change. This emphasis on performance assessment stems from three sources: (1) a backlash against the pressure for accountability through standardized testing, (2) the expansion of cognitive science (with its emphasis on constructivist teaching and learning), and (3) concern from the business community that schools are not adequately preparing youth for today's workplace. In addition, several national, nongovernmental projects designed to address curricular, instructional and assessment reform have gained prominence in recent years. These include the New Standards Project, the Coalition of Essential Schools, and the College Board's Pacesetter program, all of which have influenced a shift toward the use of performance assessments (Khattri, Reeve, and Kane 1998).

Harris and Kerby (1997) believe that the strongest argument in favor of performance assessments is the chance that they will balance the scores of students who perform relatively poorly on multiple-choice tests. For example, men tend to outperform women on multiple-choice tests, so essay tests could yield an inappropriate misclassification of women's knowledge or abilities based on these types of test scores.

The national standards that have been developed in many curricular areas, including the technology education standards (currently being developed by the International Technology Education Association), emphasize the acquisition of higher-order thinking and process skills. Unfortunately, when these curricular reforms encounter high-stakes decisions about students, programs, or schools based on mandatory standardized tests of basic skills, teachers have little incentive to pursue alternative approaches to instruction, and "the tests win out" (Darling-Hammond, Ancess, and Falk 1995, p. 10).

One can also trace the emphasis on accountability testing in general to our enhanced *ability* to test. Increasingly sophisticated tools, from the first *Iowa Test of Basic Skills* in 1929 to our current capacity to process enormous quantities of data electronically, have contributed to a kind of technology-driven push for accountability testing. As Rothman (1995) notes, Americans, fascinated by technological solutions to social problems, find that tools such as electronic scoring and recordkeeping make testing almost irresistible. More positively, as computer software becomes more sophisticated, it is becoming more possible to analyze rich qualitative data on a large-scale basis.

Federal Mandates and Initiatives

The *Goals 2000: Educate America Act* passed in 1994 mandated that states detail how student performance will be measured against established standards. Goals 2000 provided federal funding for the development of standards-based education systems, which provide the base for authentic assessments. Other federal mandates include Title I legislation, which is designed to encourage a move away from norm-referenced testing by allowing districts the flexibility to develop their own standards and assessments, provided they are as rigorous as those of the state (Khattri, Reeve, and Kane 1998). The net result of federal legislation promoting assessment alternatives is that "substantially more assessment is likely to occur in our nation's schools and to take place in areas traditionally not assessed (such as the arts) using assessment strategies (such as performance assessments and portfolios) not typically used" (Roeber 1997, p. 6).

The Carl D. Perkins Act requires states to develop performance standards for vocational programs. The states are also required to measure the effectiveness of vocational education programs related to the attainment of identified skills, school retention, program completion, job placement, and the progress of special populations. The 1990 Perkins legislation marked a "significant turning point in federal accountability by explicitly tying the process of state and local review to standards based on outcomes" (Office of Technology Assessment 1994, p. 5). As a

result, state assessment activity in vocational education increased during the 1990s to meet these accountability requirements. The Perkins Act does not, however, specify the types of assessment strategies that must be used.

In general, vocational skills assessment falls into four categories: academic skills, job-specific vocational skills, generic workplace skills, and broad technical skills. The diversity of assessment methods used to measure these skills is broad, ranging from student portfolios, to structured ratings of student capabilities demonstrated through classroom work, and organized competitive events (Office of Technology Assessment 1994). According to an OTA survey of state assessment directors, it is in the areas of vocational skills and generic workplace skills that the greatest expansion of assessment activities is likely to occur. In fact, vocational educators have used authentic assessment strategies and tools for many years.

Another federal initiative that has spurred the move toward performance-based assessment is the U.S. Department of Labor *Secretary's Commission on Achieving Necessary Skills* (SCANS) Report. "The SCANS commission envisioned setting proficiency levels for SCANS competencies and developing an associated assessment system based on demonstrating SCANS competencies through applied, contextualized problems" (Khattri, Reeve, and Kane 1998, p. 5).

The effectiveness of using testing to implement educational standards and ensure accountability for outcomes is yet to be determined. Although there continues to be considerable political and popular support for the concept of accountability through standards and assessment, significant technical, political, and logistical problems remain (Madaus and O'Dwyer 1999; Milne 1998; Wildavsky 1999).

Statewide Efforts to Use Authentic Assessment

Statewide systems of standards and measures of performance were mandated by the 1990 Perkins Act, which required an accountability system built around standards, outcomes, and performance measures. These systems were required to address mastery of academic and occupational skills, program completion, and employment. The framework of standards and measures adopted by each state should serve as a common tool for evaluating and improving vocational education programs (Milne 1998).

Data from the 1992 National Assessment of Vocational Education (NAVE) Omnibus Survey showed that virtually all states were in the process of developing performance standards, and over 75% of states were assessing (or were planning to assess) secondary student performance based on these standards. Prior to 1991-92, only 18 percent of the states were involved in this type of aligned standards-based process (Milne 1998). It should be noted that these data do not indicate what type of assessment is being used or planned.

There are a variety of performance measures that can be adopted to assess vocational programs. These can be grouped into the following categories: enrollment numbers, academic skills, occupational skills, school completion, job placement,

wages, and/or job retention. Of these, several can appropriately be considered forms of performance assessment. In the NAVE survey, more than 80 percent of the states reported plans to use at least five of these seven types of measures (Stecher, Farris, and Hamilton 1998). The most significant trend was toward a greater use of skill measures, particularly those involving advanced skill measures.

State vocational education officials, school administrators, and local vocational education administrators and staff have been instrumental in developing standards and measures for vocational education in more than 70 percent of the states surveyed in the NAVE study. Representatives from special populations were the only other group that has been heavily involved in the process. Employers, students and parents were also consulted in approximately 85 percent of the states (Stecher, Farris, and Hamilton 1998). Thus, the primary stakeholders in vocational education have been engaged in this developmental effort.

The Perkins Act has included provisions permitting states to adjust state performance standards to accommodate special populations, school resources, and local conditions. Fifty percent of the states report that they plan to make adjustments accordingly. In 64 percent of states, all students who take vocational courses were measured, whereas only 7 percent apply performance measures to the most narrowly defined population (e.g., vocational completers). Interestingly, those states that have done the most to promote academic-vocational integration were found to have also done significantly more with performance assessment measures (Stecher, Farris, and Hamilton 1998). This is good for the academic areas and it also reflects positively on vocational education.

A case study of 16 districts and states conducted by Khattri, Reeve, and Kane (1998) found that the characteristics of performance assessments varied from site to site. There is a wide range in the type and complexity of tasks required of students under the umbrella of performance assessment. Everything from open-ended, short-answer questions to completion of extended projects can make up the universe of "constructed responses," which are a characteristic part of performance assessments. Significant differences also exist between state testing policies for general education and vocational education. What is most important—

> no state vocational education agencies directly administer a program of mass testing or assessment of all students at a fixed point in time. In most states, the primary assessment responsibility of the agency is to set policies for local programs to follow. (Office of Technology Assessment 1994, p. 11)

According to the OTA, there is virtually no tradition in vocational education for the use of norm-referenced tests. Vocational education has long embraced the concepts of competency assessment, skill attainment, active student involvement, and assessment embedded within instruction. "In all of these respects, the traditions of testing and assessment in vocational education resemble what is [now] being advocated elsewhere in the rest of education" (ibid., p. 40).

California, Kentucky, Maryland, and Vermont are considered to be early leaders in the development and use of state-level performance assessments (Khattri, Reeve, and Kane 1998; Rothman 1995). California's Learning Assessment System (CLAS) was canceled by Governor Pete Wilson in 1994 following strong public criticism, but the programs in the remaining three states appear to be fairly well established. The Kentucky Instructional Results Information System (KIRIS), in full-scale implementation since the mid-1990s, includes both multiple choice and performance event components in assessment of vocational studies and "practical living," as well as art, math, reading, science, and social studies (Khattri, Reeve, and Kane 1998).

The OTA survey conducted in 1994 found that states increasingly appear to be expanding their use of written testing in vocational education "at the very time that questions are being raised in the rest of education about the effectiveness of standardized testing" (p. 59). This raises the question of what the long-term effects on instruction in vocational education are likely to be.

Implementation Issues

The increased development of performance measures, reported in the 1992 National Assessment of Vocational Education surveys, led to an increased burden on state officials. Vocational education staff in almost 80 percent of the states reported having more responsibilities related to these tasks than they had a decade earlier (Stecher, Farris, and Hamilton 1998). These researchers suggest that the expertise of educators at the local district level should be tapped when establishing new assessment systems, thus saving time and effort. This makes sense, not only from a resource standpoint, but also in terms of commitment and expertise.

Additionally, states can look to national projects or commercial entities for assistance in developing, delivering, and/or scoring alternative testing systems. Performance assessments have always been a part of some of the College Board's Advanced Placement (AP) programs. For example, the Studio Art Portfolio Evaluation has no written or multiple-choice portions. The AP Art Portfolio is one prominent example of an established, national portfolio examination (Khattri, Reeve, and Kane 1998).

In Kentucky, students are expected to complete performance tasks and submit a portfolio of best work over the course of a school year, in addition to a more traditional component to its statewide testing system (KIRIS). These state assessments, with the exception of the student portfolios, are scored by a private firm called Advanced Systems in Measurement. Maryland, which has also incorporated traditional and performance components into its state-level assessments, contracts with CTB Macmillan/McGraw-Hill to operate its program (Rothman 1995). This use of the private sector to develop and implement large-scale assessments is probably wise, given the demands on human and capital resources required to conduct large-scale performance assessments.

Obstacles and Challenges

The obstacles or challenges facing states that want to implement performance assessments have been grouped into two broad categories: practical challenges and technical challenges (Roeber 1997). Practical challenges are those that involve administering and scoring large-scale assessments, whereas technical challenges have to do with the validity and reliability of the assessments themselves. These challenges are present in any assessment situation. They are particularly problematic when assessment is conducted on a large-scale basis. Both types are discussed further in this section.

A primary problem with all large-scale assessment is accurately matching the performance being tested with the stated goals and objectives. Mager (1973) provides a humorous example to illustrate the problem. Suppose an instructor gave you the objective "On a level paved street, be able to ride a unicycle 100 yards without falling off." You work hard to develop this psychomotor skill, only to find out on assessment day that the "test" consists of the following questions:

> Define unicycle.
> Write a short essay on the history of the unicycle.
> Name at least six parts of the unicycle. (p. 1)

The assessment items being used in this example suffer from a lack of *validity*. The obvious message is that educators must be careful to match the assessment to the desired behavior or condition, a task not always easy to accomplish. The technical terms for this are *construct* and *content validity*.

Sources of Invalidity

Messick (1996) describes two major threats to validity of performance assessments. *Construct underrepresentation* means that the assessment is too narrowly focused, failing to include important dimensions of the knowledge or skill it aims to assess. *Construct-irrelevant variance* refers to assessments that ask for responses that are not relevant to measuring the desired knowledge and skills. Thus, "a primary validation concern is the extent to which the same assessment might underrepresent the focal construct while simultaneously contaminating the scores with construct-irrelevant variance" (Messick 1996, p. 5). If the irrelevant tasks are overly difficult, assessment scores will likely be invalidly low. If the irrelevant tasks are overly easy, assessment scores may be invalidly high.

Crocker (1997) provides an excellent overview of the elements that must be considered when judging content representativeness. These include "(a) the *relevance* of the test item content to the knowledge domain of interest and (b) the *balance of coverage* of the items in relation to the breadth of the domain. Some experts also consider review of the *technical quality* of items and *fairness* to examinee subgroups" (p. 84). The problem of subjective decision making with regard to test item content is exacerbated on performance assessments because they have

fewer, and thus more heavily weighted, items. Subjectivity can also be introduced through scoring rubrics, which are influenced by the preferences of the rubric developers (Crocker 1997).

A concrete example helps to illustrate the issue of content representativeness. A student has mastered 90 out of 100 concepts from the material to be tested. Given time limitations, not all concepts will be included on the test. With a multiple-choice test, there could be 60 items, compared to 6 items on a constructed-response (essay) test. The likelihood of achieving good content representativeness is much higher on the test with 60 discrete items than the test with 6. Theoretically, all 6 essays could come from the 90 known concepts or from the 10 unknown concepts. So the students' scores could range from 0 to 100 percent. In these circumstances, "the essay exam is much more likely to underestimate or overestimate the student's true knowledge of the domain and result in an erroneous decision about the student's competence." In large-scale testing programs, there is no opportunity to mitigate this possibility with a variety of other classroom scores (Phillips 1993, p. 108).

Another problematic dimension of validity has to do with the consequences associated with interpreting scores. Specifically, the concern is that any negative impact that results from the use of an assessment should not stem from any source of test invalidity. Assessment must include efforts to discover the intended and unintended consequences, in the short- and long-term, of how scores are used (Messick 1996). In high stakes, large-scale assessment efforts, these concerns are even greater. Relatively minor technical challenges can escalate into major political issues (Barton 1999; Wildavsky 1999).

Because of the particular, and sometimes sweeping, claims made for the benefits of performance assessments, there are some specialized criteria by which they should be judged. To be worthwhile and motivational educational experiences in their own right (as the claims go), the tasks posed should be meaningful to students and clearly communicate what is expected. These traits have been referred to as "meaningfulness" and "transparency" (Dunbar, Koretz, and Hoover 1991; Messick 1996, p. 13).

Yet another validity issue has to do with the variability that can be introduced by using different methods or prompts to introduce the performance task. Student performance can be "extremely sensitive to subtle changes in format and presentation," resulting in scores that do not truly reflect ability levels (Phillips 1993, p. 11).

Fairness

Bond, Moss, and Carr (1996) identify two aspects of fairness with regard to assessment. The first is test *bias*, which relates to the validity of an interpretation or action based on test scores. Bias exists when there is evidence of differential validity for any relevant subgroup of persons assessed. The second aspect of fairness relates to the soundness of the educational system upon which the assess-

ment is based, or what these authors call *equity*. In other words, did all students being assessed have access to the same quality of educational experiences? Evidence suggests that curricular changes caused by traditional standardized testing have affected nonwhite students disproportionately. They are more likely to spend time in direct test preparation (content drilling rather than more motivational types of activities) than their white counterparts (Bond, Moss, and Carr 1996). Given the differential access to high-quality education, there is no reason to expect that underserved minority groups will fare any better with performance assessments than they have traditionally done, and may actually do worse (Herman, Klein, Heath, and Wakai 1994). Studies of performance assessments in science found considerable variance in mean performance from one ethnic group to another (Dunbar, Koretz, and Hoover 1991).

Reliability

Reliability refers to the consistency of a measure over time, closely related to the concepts of generalizability and comparability described later. Among the many challenges to the appropriateness of performance assessment scores, interrater reliability is of least concern. Evidence from several studies indicates that high levels of agreement between scorers can be achieved, given sufficient training. The issues of where to set cut-off scores and how to deal with scores that fall near those cut-off points, however, have not yet been adequately addressed (Jaeger, Mullis, Bourque, and Shakrani 1996; Khattri, Reeve, and Kane 1998; Shavelson, Baxter, and Gao 1993).

Generalizability

Concerns about the content representativeness of performance assessments seek to ensure that interpretation of test scores need not be limited to the sample of assessed tasks, but rather be *generalizable* to the broader set of skills and abilities desired (Yen 1993).

> [The] issue of generalizability of score inferences across tasks and contexts goes to the very heart of score meaning. Indeed, setting the boundaries of score meaning is precisely what generalizability evidence is meant to address. However, because of the extensive time required for the typical performance task, there is a conflict in performance assessment between time-intensive depth of examination and the breadth of domain coverage needed for generalizability of construct interpretation. (Messick 1996, p. 11)

One way this has been addressed has been through the use of "matrix-sampling," where different samples of students perform different (but only a few) sets of tasks. The amount of time spent by any one student is minimized. Scores are evaluated in the aggregate, permitting comparisons between larger groups such as districts, states, or nations, rather than at the individual student level. This makes matrix-sampling useful for large-scale efforts like the National Assessment of Educational Progress (Calderone, King, and Horkay 1997).

In addition to matrix sampling, Brennan (1996) suggests combining performance assessments with traditional testing. Alternatively, a series of short-term performance tasks could be devised, so that a larger number of items could be included. In this way, the benefits of performance assessment (such as greater authenticity) could be realized without sacrificing generalizability.

Comparability

Another goal of large-scale testing is that scores should be comparable over time as well as from sample to sample. This comparability requires that the content assessed with each sample remain proximally the same. On multiple choice tests the number of items means that the significance of any one item is small and it is easy to create comparable tests over time. Because performance tasks are fewer in number and are more distinctive, they must be sampled with greater care than multiple-choice items, so that the knowledge and skills assessed remain stable over time (Haertel and Linn 1996).

The Challenges of Setting Assessment Standards

Closely related to comparability is the issue of setting assessment standards against which student work is to be judged. Jaeger et al. (1996) identify some factors that complicate the process of setting standards. One is that performance standards rarely occur naturally in ways that make it obvious where the boundary between acceptable and unacceptable work lies. This is less true for some skills tasks, such as might be found in vocational areas. Students can either perform the task correctly, or they can't. Another factor is that the people who set performance standards are not always trustworthy judges of the quality of the standards they have set. If performance assessments are to be used in accountability-based strategies for promoting systemic educational reform, the issue of setting assessment standards must be addressed.

Developing comparable standards across performance assessments appears to be the most problematic venture of all. There is, first, the problem of designing appropriate performance tasks in terms of content standards, and identifying student work on the tasks that exemplifies success in meeting the standards. Implementing such assessments under consistent conditions and evaluating resulting performances reliably pose enormous operational challenges. (ibid., p. 87)

Curriculum standards specifying what students must know and be able to do as a result of instruction have been developed for mathematics, English, civics, geography, history, foreign languages, science, social studies, and the arts. In Spring 2000, standards will also be issued for technology education. Formal assessment standards for determining when standards have been met are not typically a part of these documents (Jaeger et al. 1996). However, some of these national standards projects are in the process of developing (or proposing to develop) assessment standards (e.g., the Technology for All Americans project).

The essential concern with setting cut-points for levels of performance is that the process imposes artificial dichotomies on what is in reality a continuum of proficiency. In the real world, proficiency does not occur at discrete, easily recognizable points. "The problem is how to treat the gray areas around the cut-points, since a certain proportion of examinees just above or just below a cut-point will almost inevitably be misclassified due to measurement error" (ibid., p. 104). One study found that nearly six times as many students would fail an assessment when using one standard-setting measure as opposed to another. As with other assessment considerations, political realities demand that issues surrounding interpretation of scores be done in a manner that accounts for the many "shades of gray" that are involved in making assessment-based decisions.

A related concern is the degree to which curriculum standards, once adopted by a state, are implemented at the local level. As a result of its Education Reform Act of 1993, Massachusetts developed a series of tests for grades 4, 8, and 10, based on its curriculum framework for technology education. Results from the first year of testing show that student performance at the higher grade levels is significantly lower than at grade 4. One reason suggested for the low first-year test scores at grades 8 and 10 is that, at the time the tests were administered, only 30 percent of the school districts in the state had aligned their curriculum to the new standards (Bouvier and Corley 1999).

Multiple Purposes

The primary purposes of assessment are to monitor student progress, establish accountability, certify student achievement, and align curriculum, instruction, and assessment. The ability to judge the effectiveness of performance assessment systems, particularly at the state level, is hampered by the fact that many systems are set up to achieve multiple purposes. Factors that facilitate the achievement of one purpose (e.g., standardization for purposes of accountability) may serve as barriers to the achievement of another purpose (e.g., informing instructional practice) (Khattri, Reeve, and Kane 1998). With large-scale assessment, there is an increased chance that the challenges associated with competing priorities and goals will occur.

Cost

Developing and implementing performance assessments is an expensive undertaking. Other aspects of the assessment reform process that require financial support include research on assessment methodology, delivering professional development, disseminating information, storage space for assessment materials, time spent by teachers preparing for assessments, and more (Khattri, Reeve, and Kane 1998). Additionally, some of the costs of performance assessment are not obvious or are not known, such as time spent with governmental and nongovernmental agencies, state departments of education, etc. (Hardy 1995).

Estimates of costs are also difficult to establish because there are few authentic assessments in place on a large scale (Rothman 1995). In 1992, the OTA reported

on testing in American schools. In this report, the costs associated with traditional testing in large districts (including both direct and indirect costs) were estimated to be approximately $37 per student. The General Accounting Office (GAO), in a similar study, estimated per student costs of traditional testing to be much lower, at $16. Estimates of the cost for performance assessments made by these same agencies ranged from twice as much as traditional (GAO) to 3-10 times as much as traditional (OTA) (Rothman 1995).

A limited number of private companies are developing performance assessments. Hardy (1995) examined several large-scale performance assessment programs to estimate the costs of development, implementation, and scoring. He notes that development costs can often be hard to assess, particularly when existing staff is used. Based on available data, however, development costs ranged from $5,000 to over $14,000 per task. Costs tend to be lower when the student outcomes are well defined, when smaller sample sizes are used to pilot assessment tasks, and when the size of the development teams is kept to a minimum. Development costs can also vary considerably depending on the content area. When local educators are used to develop items, the cost of their training must also be added.

The costs of performance assessment fall into three categories: development, administration, and scoring. All three can be vary widely depending on the nature of the assessment task, the type of work produced, and the amount of information required from individual responses. Hardy (1995) examined several large-scale performance assessment programs to estimate the costs of development, implementation, and scoring.

Administration costs include materials and staffing. Performance assessment kits for science and mathematics tasks developed by the National Assessment of Educational Progress (NAEP), by the Educational Testing Service for the state of Georgia, and others ranged in cost from a low of $.70 to a high of $13.50 per kit. Ways to reduce the cost of materials include testing only a sample of students or using the same kit over a multiyear period and prorating its cost. Another possible approach is to require all classrooms to have a common set of equipment or materials that would be used in the classroom over the course of a school year for instruction, as well as on the performance test. Staffing costs for actual delivery of the assessments can also be difficult to calculate, particularly when local personnel are used. Kentucky uses external task administrators at a cost of approximately $5 per student (Hardy 1995).

The costs of scoring performance assessment tasks are considerably higher than those associated with scoring traditional multiple-choice tests. Most performance tasks require some form of human analysis, if not outright hand scoring. Estimates of scoring costs, largely based on writing tasks conducted in various states, range from $3-$6 per student (Hardy 1995).

The U.S. Office of Technology Assessment estimates the cost of using performance assessments will be from 3 to 10 times greater than the costs associated with traditional tests. Other estimates have suggested they could be up to 60

times more costly. However, as performance assessments are more widely used, their cost per assessment unit will likely decrease (Hardy 1995). The savings may be less significant because per student development costs per student drop with larger numbers, whereas the other major cost factors (e.g., materials, administration, scoring) do not (Stecher 1995).

Professional development in the use and scoring of authentic assessments is critical to their success (Khattri, Reeve, and Kane 1998). The cost for the level of training required for reliable scoring of these assessments is considerable, particularly compared to traditional forms of testing. Other aspects of the assessment reform process that require financial support include research on assessment methodology, disseminating information, storage space for assessment materials, time spent by teachers preparing for assessments, and more (ibid.). Some of the costs of performance assessment are not obvious or are not known, such as time spent with governmental and non-governmental agencies, state departments of education, etc. (Hardy 1995).

According to the NAVE 1992 Omnibus Surveys, less than half of school districts surveyed reported any increase in state assistance with accountability assessments, and fewer than 20 percent noted any state-sponsored training programs on student assessment or performance assessment for vocational educators (Stecher, Farris, and Hamilton 1998). In addition, although their use was mandated by the 1990 Perkins Act, few districts used Perkins Title II basic grant funds to develop or expand vocational performance assessment systems.

Stecher (1995) calculated the approximate cost of traditional (paper and pencil) testing in science using the California Test of Basic Skills at $.30 per student. By contrast, open-ended written-response items on the same test cost $4.80 per student per prompt. In his study, Stecher examined the costs of developing, implementing, and scoring performance tasks for science. The study suggests that the cost of hands-on science assessment can run as much as 100 times higher than standardized multiple-choice tests. Hands-on science performance tasks developed and implemented in this study were calculated at $30 per student per test period of 45-50 minutes, provided 100,000 students take the test and the economy of scale is realized. With fewer students taking the test, costs will go up significantly. This cost does not include teacher time for administering the performance test, which was considered "contributed time."

Time

There are two dimensions of time that may be problematic when using large-scale performance assessments. One challenge is the amount of student testing time required to administer a sufficient number of items to satisfy validity and generalizability concerns. The second challenge is the turn-around time for test results, which can sometimes be as long as several months. "This time lag between assessment and reporting is so large that local educators may view the results (and the overall assessment program) as relatively useless" (Roeber 1997, p. 8).

Preliminary Reactions to the Large-Scale Use of Authentic Assessment

Some responses to large-scale adoption of performance assessments have been less than positive. Problems that face all large-scale assessments of any type, including inappropriate testing practices, breaches in test security, adverse impacts on historically disadvantaged groups, and others can plague performance assessments just as they do traditional multiple-choice tests. In fact, some preliminary data suggest that many of these issues may be even more pronounced with performance assessments (Phillips 1993). Implementation in some states has led to poor results, whereas other states continue to plan to implement performance assessments in the near future. Givens (1997) suggests that communication regarding the "myriad problems" associated with this form of testing has to date been limited.

In some cased, the issue has become highly political. For example, in the early 1990s, educators in Littleton, Colorado attempted a system of reforms that they hoped would help students develop better problem-solving and communication skills. They established standards, redesigned instructional practices, and created new performance assessments. Although many teachers, parents, and students felt positive about the reforms, a vocal and, as it turned out, powerful group of residents opposed them. They viewed the new assessments as being too new and untried, and too reliant on teachers' judgments to be appropriate for high-stakes decisions such as determining whether students would graduate from high school. The critics also complained that the schools should not focus on problem-solving abilities, but rather on knowledge of a core body of information. In 1993, in a heated school board election, three community members who opposed the changes were voted into office. They subsequently scrapped the reform program and removed the superintendent of schools from office (Rothman 1995).

Rothman believes there are several reasons why reforms based on standards and new assessments have met with strong resistance in some communities. First, the establishment of explicit standards, while necessary from the standpoint of clearly communicating expectations, also invites challenges about what we really *do* want students to know and be like, and who should decide that. Second, many opponents object to the methods of teaching, constructivist in nature, that shift greater responsibility for acquiring knowledge onto the student. Critics of California's CLAS reforms raised similar objections, saying the assessments chosen were designed to measure attitudes and beliefs, rather than academic knowledge and skills (Lewis 1996; Rothman 1995). As noted earlier, use of the CLAS system was halted.

How Are Data Being Used?

At a 1993 meeting of the International Congress on School Effectiveness, educators from the United States, the United Kingdom, Sweden, and Holland discussed the role assessment plays in school reform. A widely held belief that emerged from

these discussions was that "school improvement would not occur if schools were left to take action on their own." In the absence of external evaluations, participants agreed, schools would continue to do what they had been doing (Riley and Nuttall 1994, p. 126). However, if alternative assessments are to drive education reform, as many would like, the data should at some point be fed back into the decision-making structure at the local level, where school improvement must be sustained.

The OTA (1994) found that states use occupational skill assessment data differently than they use data from academic skill assessments. Occupational data is most often used to evaluate student attainment for certification or program completion. The second most frequent use is for accountability (is the program doing what it is supposed to be doing?). The third most frequent use is for making decisions about the improvement of courses, programs, or schools. In other words, schools are least likely to use assessment information to improve programs. They are unlikely to link information about academic skills to instruction, but rather -- collect that information for Perkins accountability purposes only.

The practice of "teaching to the test" is a recognized outcome of high-stakes assessment. Teaching to the test can cover a range of interventions, not all of which are ethical. One problem with teaching to the test is that it can narrow the curricular focus to only what is on the test, or sacrifice material at the expense of covering tested material. Another problematic trend related to standardized tests is that disadvantaged students are less likely to receive instruction in science, art, and thinking skills, and more likely to receive drilling on the so-called basic skills (Rothman 1995). The reality of high-stakes testing is that it *will* have an effect on instructional practices. For this reason it is imperative that teachers have a clear understanding about the measures being taken, so that they can organize their instruction accordingly (Popham 1999).

Certain educators have questioned the trend toward large-scale performance assessment and, in fact, the whole foundation upon which large-scale assessments of *any* kind are based (Andrews 1997; Barton 1999; Haertel 1999; Lewis 1996; Lissitz 1997; Madaus and O'Dwyer 1999). Lissitz (1997) believes there is little evidence that performance assessment will lead to better teaching, any more than traditional assessments have done. He and other critics maintain that, if we really hope to reform classroom teaching, we should advocate for change in the teaching environment, not for changes or additions to state-level accountability testing. According to Eisner, what really needs to change is the conception of schools in the minds of the public. "A shift needs to be made from a conception of schooling as a horse race or a kind of educational Olympics to a conception of schools as places that foster students' distinctive talents" (Eisner 1999, p. 660).

At the bottom line, the question that must be addressed is: "What will performance testing do that cannot be accomplished more reliably, quickly, and cheaply with fixed-response (multiple-choice) instruments?" (Harris and Kerby 1997, p. 132). Implicit in this question is the recognition that large-scale accountability testing is a political imperative. Given this, the challenge is to identify those

situations for which performance testing represents the most valid and appropriate form of assessment.

Viewed from a different perspective, the question might be "what are the ramifications of *not* including performance components in large-scale testing in vocational education"? In the Massachusetts Technology Education assessment, the development team concluded that all but 1 percent of the standards could be suitably evaluated using a large-scale written assessment. Faced with the argument that authentic assessments might better reflect the hands-on nature of technology education (and thus result in better test scores), one state official responded that the "development committee will continue to explore, identify, and evaluate content that can be included in a written model" (Bouvier and Corley 1999, p. 29). This suggests, perhaps unintentionally, that the nature of the curriculum could change to accommodate traditional modes of testing.

Successful Models of Implementation

In spite of the problems, there have been some successful implementation models. The National Assessment of Educational Progress (NAEP), conducted by the U.S. Department of Education for over 29 years, provides one model for carrying out large-scale performance assessments. NAEP tests were first administered on a statewide basis in 1990, and since that time most states have voluntarily participated. In 1997, the NAEP included a small-scale operational assessment of performance in the visual and performing arts, in addition to the main assessment areas of reading, writing and civics. Recent NAEP tests reflect the trend toward authentic assessment that was begun in 1992. Reading and writing test items include a large proportion of constructed-response questions. The civics assessment items also reflect this trend. The operational arts assessments used in 1997 required the students to create, perform, and/or interpret works within the discipline (i.e., art, music, theater, or dance). Student "responses" were recorded via videotape, audiotape, or photograph (Calderone, King, and Horkay 1997).

The sheer numbers of test items and scorers needed to process student responses is daunting, and the process used provides valuable insights into how large-scale performance assessments should be carried out. For example, in the 1996 NAEP, nearly 9 *million* constructed responses in mathematics and science were scored by a total of 675 scorers, with an elapsed scoring time of only 12.5 weeks (ibid.). A high level of reliability in scoring was achieved through the following steps:

- The development of focused, explicit scoring guides that match the assessment frameworks;
- Recruitment and rigorous training of qualified scorers, including post-training qualifying tests;
- The use of a digital image processing and scoring system that allows all responses to a particular exercise to be scored continuously until done, thus enhancing validity and reliability of scorer judgments;

- Monitoring scorer consistency by "backreading" approximately 10 percent of each scorer's ratings, and calibrating scores to be sure that scorer drift (the tendency to grade an item higher or lower over time) is minimized;
- Checking for interrater reliability to ensure consistent ratings; and
- Keeping careful documentation of the entire process.

Historically, vocational educators have relied on performance assessments at the classroom level. Four vendors have created assessment tools for vocational education on a national level. Although their use and influence remain relatively small, they provide information regarding the trends in assessment on the national level for vocational education.

Work Keys. Work Keys is a system developed by ACT for teaching employability skills and generic workplace skills. All of the Work Keys tests emphasize workplace application of skills rather than academic applications. Work Keys materials include tests suitable for large-scale, high-stakes testing, along with other reporting tools.

V-TECS. The Vocational-Technical Consortium of States (V-TECS), founded in 1973, has as its goal the promotion of competency-based vocational education. Beginning in 1986, V-TECS created banks of test items for members to use in constructing their own competency-based tests. The test banks include both written and performance-based items. The V-TECS materials are readily available and frequently modified to fit local needs, and thus do not represent secure tools for large-scale assessments (Office of Technology Assessment 1994).

NOCTI. The National Occupational Competency Testing Institute (NOCTI) began developing competency tests for vocational students in the late 1970s. Since that time the organization has created, with its member states, over 70 Student Occupational Competency Achievement Testing (SOCAT) exams. The SOCAT tests have both a written and a performance component, tied to the competencies required of entry-level workers in the respective fields for which tests have been developed. Performance tests are supposed to be judged by industry representatives, who examine both the process and the product. "Although NOCTI has traditionally discouraged the use of the written tests alone, in 1992 the organization began making the written test available for pretesting because of accelerated interest in using it to fulfill Perkins requirements" (ibid., p. 79).

C-TAP. A program known as the Career-Technical Assessment Project (C-TAP) was developed for the state of California by the Far West Laboratory for Education Research and Development. Within occupational clusters, students will be certified job-ready through a series of cumulative and administered assessments. These cumulative assessments include supervised practical experience, an assessment project, and a portfolio of work. The administered assessments consist of structured exercises given to students at a certain time, and include project presentations, written scenarios focusing on solving a technical problem within the vocational area, and an on-demand test (ibid.).

Suggestions for Implementation

"The adequacy of the amount of time allowed for development, introduction, and institutionalization of assessment reform can have a dramatic impact on a state's ability to sustain its reform efforts" (Khattri, Reeve, and Kane 1998, p. 74). Unfortunately, when performance assessment measures are introduced into the political realm, the pressure to show quick results is greater. A low level of involvement on the part of teachers in the development and implementation processes does impede the acceptance of changes in teaching practice (ibid.).

Several policy implications for adoption of large-scale performance assessments have been identified by Khattri, Reeve, and Kane (1998):

Clearly state the primary purpose of the assessment system.
Match the format of the system with the purpose.
Coordinate assessment reform with other elements of education reform and with other testing requirements.
Articulate in clear and simple terms the content and performance standards the assessment system is intended to measure.
Institute procedures to ensure the technical quality and fairness of the assessment system.
Design a system that contains a mix of different types of performance assessment tasks and procedures, to obtain a comprehensive picture of student learning.
Tap existing resources when developing performance-based assessments.
Communicate to the public the purposes of, and the theory underlying, the assessment. (pp. 153-157)

Because validity issues are such a major concern with any large-scale assessment, and in particular authentic assessments, using existing resources such as the Educational Testing Service, NOCTI, and others may be the best approach for the states. As Barton (1999) notes, the use of standardized tests for accountability purposes "without meeting standard and well-known methods of validation amounts to testing malpractice" (p. 9). Professional testing organizations, which specialize in the development of assessment tools, can serve as contractors to state-level and local education agencies. NOCTI, for example, provides customized assessments for local clients, which could be tailored to address vocational education standards adopted by a state (NOCTI, n.d.). In addition, states can look to national projects or commercial entities for assistance in delivering, and/or scoring alternative testing systems. For example, in Kentucky students are expected to complete performance tasks and submit a portfolio of best work over the course of a school year, in addition to a more traditional component to its statewide testing system (KIRIS). The state assessments, with the exception of the student portfolios, are scored by a private firm called Advanced Systems in Measurement. Maryland, which also incorporates traditional and performance components in its state-level assessments, contracts with CTB Macmillan/ McGraw-Hill to operate its program (Rothman 1995).

Regardless of who is responsible for development of assessment tools, a diverse panel should be assembled to develop and score the assessment, to reduce internal bias. Perhaps most critically, steps must be taken to ensure that the content framework upon which the assessment is based is appropriate (Bond, Moss, and Carr 1996). This can be accomplished, in part, by linking content to national standards, where available. Up-to-date job and task analyses, which have traditionally formed the basis for content frameworks in vocational education, can provide straightforward standards upon which to base valid performance assessments.

In a climate of education reform, unfortunately, new assessment measures are sometimes introduced in an effort to bring about curricular change, and there is political pressure to show quick results. This pressure can impede the acceptance of desired changes in teaching practice, particularly when teachers have not been involved in curriculum reform efforts, or when they are given inadequate time and training to make the necessary changes in teaching practice (Khattri, Reeve, and Kane 1998). Some educators also worry that moving too rapidly toward adoption of state performance assessments might backfire. Determining where and when both traditional and performance assessments can most effectively be used is more important than advocating for one type versus the other (O'Neil 1992).

Finally, national, state, and local assessments should be coordinated so that together they present a coherent view of student performance. In a comprehensive system, for example, various assessment strategies can be implemented. At the local level, portfolio assessment could provide data to improve instruction. At the state level, matrix sampling could be used to strengthen the local data, and could provide information for reporting purposes (Roeber 1997). State assessments can be linked with national efforts like the NAEP to provide meaningful, comparable data (Barton 1999). More than one measure should "count" if assessment data are used to make high-stakes decisions related to grade-level promotion, graduation, or teacher income. In this way, the multiple purposes of assessment can be better addressed.

Summary

Performance assessments are viewed by many as a key component of assessment reform that will, in turn, drive curricular and teaching reforms. Performance assessments can provide more authentic measures of student capability than standardized, multiple-choice tests, while at the same time encourage instructional practices that emphasize acquisition of more sophisticated thinking and process skills.

Vocational education has a long history of using criterion-referenced, standards-based measures and performance assessments. With some exceptions, these have not been used at the state or national level. What has occurred in the current climate is that accountability measures imposed by federal Perkins funding have resulted in an expanded use of written, standardized tests in vocational education.

So far, measures of academic skills have remained largely separate from vocational assessment efforts. Studies show that use of large-scale assessment data to improve instruction is, in reality, a relatively low priority.

Implementation of performance assessment measures on a large scale carries with it a host of practical and technical challenges, including issues relating to validity, generalizability of data, cost, and equity. These obstacles have led some to suggest that performance assessment measures are best taken at the classroom level, where they can provide meaningful information for use in improving instruction. Others advocate for partnerships between schools and private test-development entities, which may be better able to solve the challenges inherent in large-scale authentic assessment.

Large-scale authentic assessment tools have the potential to be useful, particularly for vocational fields where occupational skill attainment standards can be clearly identified. Proponents of this approach must garner the political support needed for adoption of more costly authentic assessments. Decision makers must also address some significant challenges before implementing high-stakes performance measures for vocational education on a large scale. If the decision is made to adopt such measures, care must be taken at all steps to ensure that the outcome achieves the stated purposes.

References

Items with ED numbers may be ordered from the ERIC Document Reproduction Service, 7420 Fullerton Road, Suite 110, Springfield, VA 22153-2852; 703/440-1400; toll-free (United States and Canada): 800/443-3742; fax: 703/440-1408; e-mail: service@edrs.com; http://edrs.com

Andrew, R. "The Statewide Waste and Discouragement of Performance Assessment." ***Contemporary Education*** 69, no. 1 (Fall 1997): 11-14. (ERIC No. EJ 558 969)

Archbald, D. A., and Newmann, F. M. ***Beyond Standardized Testing: Assessing Authentic Achievement in Secondary School.*** Madison: National Center on Effective Secondary Schools, University of Wisconsin, and Washington, DC: National Association of Secondary School Principals, 1988. (ERIC Document Reproduction Service No. ED 301 587)

Barell, J. "Like an Incredibly Hard Algebra Problem: Teaching for Metacognition." In ***If Minds Matter: A Foreword to the Future, vol. I***, edited by A. L. Costa, J. A. Bellanca, and R. Fogarty, pp. 257-266. Palatine, IL: IRI/Skylight Publishing Inc., 1992. (ERIC Document Reproduction Service No. ED 379 070)

Barton, P. ***Too Much Testing of the Wrong Kind; Too Little of the Right Kind in K-12 Education.*** Princeton, NJ: Policy Information Center, Educational Testing Service, 1999. (ERIC Document Reproduction Service No. ED 430 052)

Bond, L.; Moss, P.; and Carr, P. "Fairness in Large-scale Performance Assessment." In ***Technical Issues in Large-scale Performance Assessment, NCES 96-802,*** edited by G. W. Phillips, pp. 117-140. Washington, DC: National Center for Education Statistics, Office of Educational Research and Improvement, U.S. Department of Education, 1996. (ERIC Document Reproduction Service No. ED 399 300)

Bottoms, G.; Pucel, D. J.; and Phillips, I. ***Designing Challenging Vocational Courses.*** Atlanta, GA: Southern Regional Education Board, 1997.

Bouvier, D., and Corley, C. "Assessing Technology Education in Massachusetts." ***Tech Directions*** 59, no. 1 (1999): 28-29.

Brennan, R. L. "Generalizability of Performance Assessments." In ***Technical Issues in Large-scale Performance Assessment, NCES 96-802,*** edited by G. W. Phillips, pp. 19-58. Washington, DC: National Center for Education Statistics, Office of Educational Research and Improvement, U.S. Department of Education, 1996. (ERIC Document Reproduction Service No. ED 399 300)

References

Brown, J. S.; Collins, A.; and, Duguid, P. "Situated Cognition and the Culture of Learning." ***Educational Researcher*** 18, no. 1 (January-February 1989): 32-42. (ERIC No. EJ 386 603)

Bruner, J. ***Toward a Theory of Instruction.*** Cambridge, MA: Harvard University Press, 1966.

Burke, K. A., ed. ***Authentic Assessment: A Collection***. Palatine, IL: IRI/Skylight Publishing, 1992. (ERIC Document Reproduction Service No. ED 425 194)

Burke, K. ***The Mindful School: How to Assess Authentic Learning. 2d ed.*** Palatine, IL: IRI/Skylight Publishing, 1994. (ERIC Document Reproduction Service No. ED 425 196)

Butts, P. "Finding Alternative Assessment Resources on the Web." ***Technology Connection*** 4, no. 7 (December 1997): 10-14. (ERIC No. EJ 555 832)

Calderone, J.; King, L. M.; and Horkay, N. eds. ***The NAEP Guide: A Description of the Content and Methods of the 1997 and 1998 Assessments. NCES 97-990, rev. ed.*** Washington, DC: National Assessment of Educational Progress, U.S. Department of Education, 1997. (ERIC Document Reproduction Service No. ED 412 294)

California Assessment Collaborative. ***Charting the Course toward Instructionally Sound Assessment: A Report of the Alternative Assessment Pilot Project.*** Sacramento: California Assessment Collaborative, 1993. (ERIC Document Reproduction Service No. ED 373 098)

California Department of Education and Far West Laboratory. ***Career-Technical Assessment Program: Teacher Guidebook***. Sacramento: Career-Vocational Division, California Department of Education; San Francisco: Far Web Laboratory, 1993.

Crehan, K. D. "An Investigation of the Validity of Locally Developed Performance Measures in a School Assessment Program." Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, IL, 1997. (ERIC Document Reproduction Service No. ED 410 275)

Crocker, L. "Assessing Content Representativeness of Performance Assessment Exercises." ***Applied Measurement in Education*** 10, no. 1(1997): 83-95. (ERIC No. EJ 538 592)

Cronin, J. F. "Four Misconceptions about Authentic Learning." ***Educational Leadership*** 50, no. 7 (April 1993): 78-80. (ERIC No. EJ 461 143)

Custer, R. L. "Rubrics: Authentic Assessment Tool for Technology Education." ***Technology Teacher*** 55, no. 4 (1996): 27-37. (Reprinted in ***The ASCD Curriculum Handbook, Section 10A: Technology Education***, edited by J. Walter, pp. 10.i-10.70. Alexandria, VA: Association for Supervision and Curriculum Development, 1996.

D'Urso, J. H. "What Happens When Students Take Part in Their Own Assessment." ***Teaching and Change*** 4, no. 1 (Fall 1996): 5-19. (ERIC No. EJ 536 816)

Darling-Hammond, L.; Ancess, J.; and Falk, B. ***Authentic Assessment in Action.*** New York: Teachers College Press, 1995.

Detterman, D. K. "Case against Transfer." In ***Transfer on Trial***, edited by D. K. Detterman and R. J. Sternberg,. Norwood, NJ: Ablex, 1993.

Dewey, John. ***John Dewey on Education; Selected Writings,*** edited by Reginald D. Archambault. Chicago: University of Chicago Press, 1974.

Dunbar, S.; Koretz, D.; and Hoover, H. "Quality Control in the Development and Use of Performance Assessments." ***Applied Measurement in Education*** 4, no. 4 (1991): 289-303. (ERIC No. EJ 446 656)

Eisner, E. "The Uses and Limits of Performance Assessment." ***Phi Delta Kappan*** 80, no. 9 (May 1999): 658-660. (ERIC No. EJ 585 676)

Falchikov, N., and Boud, D. "Student Self Assessment in Higher Education: A Meta-Analysis." ***Review of Educational Research*** 59, no. 4 (Winter 1989): 395-430. (ERIC No. EJ 412 506)

Flesher, J.W. "An Exploration of Technical Troubleshooting Expertise in Design, Manufacturing, and Repair Contexts." Ph.D. Dissertation, University of Illinois at Urbana-Champaign, 1993.

Fogarty, R. ***The Mindful School: How to Teach for Metacognitive Reflection.*** Palatine, IL: IRI/Skylight Publishing, 1994. (ERIC Document Reproduction Service No. ED 377 965)

Fogarty, R., and Bellanca, J. A. ***Patterns for Thinking: Patterns for Transfer.*** Palatine, IL: IRI/Skylight Publishing, Inc., 1987.

Fogarty, R.; Perkins. D.; and Barell, J. ***The Mindful School: How to Teach for Transfer.*** Palatine, IL: IRI/Skylight Publishing, Inc., 1992.

Frazier, D. M., and Paulson, F. L. "How Portfolios Motivate Reluctant Writers." ***Educational Leadership*** 49, no. 8 (May 1992): 62-65. (ERIC No. EJ 444 319)

References

Gillespie, C. S.; Ford, K. L.; Gillespie, R. D.; and Leavell, A. G. "Portfolio Assessment: Some Questions, Some Answers, Some Recommendations." ***Journal of Adolescent & Adult Literacy*** 36, no. 9 (March 1996): 480-491. (ERIC No. EJ 525 767)

Givens, K. "Performance Assessment Tests: a Problematic Panacea." ***Contemporary Education*** 69, no. 1 (Fall 1997): 27-29. (ERIC No. EJ 558 972)

Glaser, R. "Cognitive and Environment Perspectives on Assessing Achievement." In ***Assessment in the Service of Learning: Proceedings of the 1987 ETS Invitational Conference,*** edited by E. E. Freeman. Princeton, NJ: Educational Testing Service, 1988. (ERIC Document Reproduction Service No. ED 314 433)

Goodrich, H. "Understanding Rubrics." ***Educational Leadership*** 54, no. 4 (December 1996-January 1997): 14-17.

Haertel, E. "Performance Assessment and Education Reform." ***Phi Delta Kappan*** 80, no. 9 (May 1999): 662-666. (ERIC No. EJ 585 677)

Haertel, E. H., and Linn, R. L. "Comparability." In ***Technical Issues in Large-scale Performance Assessment, NCES 96-802,*** edited by G. W. Phillips, pp. 59-78. Washington, DC: Office of Educational Research and Improvement, U.S. Department of Education, 1996. (ERIC Document Reproduction Service No. ED 399 300)

Hamayan, E. V. "Approaches to Alternative Assessment." ***Annual Review of Applied Linguistics*** 15 (1995): 212-226. (ERIC No. EJ 501 485)

Hardy, R. A, "Examining the Costs of Performance Assessment." ***Applied Measurement in Education*** 8, no. 2 (1995): 121-134. (ERIC No. EJ 505 891)

Harris, R., and Kerby, W. "Statewide Performance Assessment as a Complement to Multiple-choice Testing in High School Economics." ***Journal of Economic Education*** 28, no. 2(Spring 1997): 122-135. (ERIC No. EJ 548 247)

Hattie, J.; Biggs, J.; and Purdie, N. "Effects of Learning Skills Interventions on Student Learning: A Meta-Analysis." ***Review of Educational Research*** 66, no. 2 (Summer 1996): 99-135. (ERIC No. EJ 533 504)

Haydel, J. B.; Oescher, J.; and Banbury, M. "Assessing Classroom Teacher's Performance Assessments." Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA, 1995. (ERIC Document Reproduction Service No. ED 392 827)

Henderson, P., and Karr-Kidwell, P. J. ***Authentic Assessment: An Extensive Literary Review and Recommendations for Administrators.*** (ERIC Document Reproduction Service No. ED 418 140)

Herman, J. L. "The State of Performance Assessments." ***School Administrator***, AASA Online. Retrieved January 13, 1999 from the World Wide Web <http://www.aasa.org/SA/dec9804.htm>

Herman, J. L.; Aschbacher, P. R.; and Winters, L. ***A Practical Guide to Alternative Assessment.*** Alexandria, VA: Association for Supervision and Curriculum Development, 1992. (ERIC Document Reproduction Service No. ED 352 389)

Herman, J.; Klein, D. C. D.; Heath, T. M.; and Wakai, S. T. ***A First Look: Are Claims for Alternative Assessment Holding Up? Project 3.2: State Accountability Models in Action.*** Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing, 1994. (ERIC Document Reproduction Service No. ED 376 213)

Hipps, J. A. "Trustworthiness and Authenticity: Alternative Ways to Judge Authentic Assessments." Paper presented at the annual meeting of the American Educational Research Association, Atlanta, GA, 1993. (ERIC Document Reproduction Service No. ED 376 195)

Huerta-Macias, A. "Alternative Assessment: Responses to Commonly Asked Questions." ***TESOL Journal*** 5 (Autumn 1995): 8-10. (ERIC No. EJ 513 672)

Jaeger, R. M.; Mullis, I. V. S.; Bourque, M. L.; and Shakrani, S. "Setting Performance Standards for Performance Assessments: Some Fundamental Issues, Current Practice, and Technical Dilemmas." In ***Technical Issues in Large-scale Performance Assessment, NCES 96-802,*** edited by G. W. Phillips, pp. 79-115. Washington, DC: National Center for Education Statistics, Office of Educational Research and Improvement, U.S. Department of Education, 1996. (ERIC Document Reproduction Service No. ED 399 300)

Jiang, Y. H.; Smith, P.; and Nichols, P. L. "Error Sources Influencing Performance Assessment Reliability or Generalizability: A Meta Analysis." Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL, 1997. (ERIC Document Reproduction Service No. ED 409 342)

Johnson, S. D. ***Knowledge and Skill Differences between Expert and Novice Technicians on Technical Troubleshooting Tasks.*** St. Paul: Training and Development Research Center, University of Minnesota, 1987. (ERIC Document Reproduction Service No. ED 284 054)

Khattri, N.; Reeve, A.; and Kane, M. ***Principles and Practices of Performance Assessment***. Mahwah, NJ: Lawrence Erlbaum Associates, 1998.

Lave, J. ***Cognition in Practice: Mind, Mathematics and Culture in Everyday Life***. Cambridge: Cambridge University Press, 1988.

Lave, J., and Wenger, E. ***Situated Learning: Legitimate Peripheral Participation.*** Cambridge: Cambridge University Press, 1991.

References

Lewis, A. C. "Making Sense of Assessment." ***School Administrator*** 53, no. 11 (December 1996): 8-12. (ERIC No. EJ 537 475)

Lissitz, R. "Statewide Performance Assessment: Continuity, Context and Concerns." ***Contemporary Education*** 69, no. 1 (Fall 1997): 15-19. (ERIC No. EJ 558 970)

Lund, J. "Authentic Assessment: Its Development and Application." ***Journal of Physical Education, Recreation and Dance*** 68, no. 7 (1997): 25-28.

Mabry L. ***Portfolios Plus: A Critical Guide to Alternative Assessment.*** Thousand Oaks, CA: Corwin Press, 1999.

Madaus, G. F., and O'Dwyer, L. M. "A Short History of Performance Assessment–Lessons Learned." ***Phi Delta Kappan*** 80, no. 9 (May 1999): 688-695. (ERIC No. EJ 585 681)

Mager, R. F. ***Measuring Instructional Intent***. Belmont, CA: Fearon Publishers, 1973.

Mager, R. F. ***Preparing Instructional Objectives. 2nd ed***. Belmont, CA: Lake Publishing Co., 1975.

Marzano, R. J.; Pickering, D.; and McTighe, J. ***Assessing Student Outcomes***. Alexandria, VA: Association for Supervision and Curriculum Development, 1993.

McDermott, J. J., ed. ***The Philosophy of John Dewey.*** Chicago: University of Chicago Press, 1981.

Messick, S. ***The Interplay of Evidence and Consequences in the Validation of Performance Assessments.*** Princeton, NJ: Educational Testing Service, 1992. (ERIC Document Reproduction Service No. ED 390 891)

Messick, S. "Validity of Performance Assessments." In ***Technical Issues in Large-scale Performance Assessment, NCES 96-802,*** edited by G. W. Phillips, pp. 1-18. Washington, DC: National Center for Education Statistics, Office of Educational Research and Improvement, U.S. Department of Education, 1996. (ERIC Document Reproduction Service No. ED 399 300)

Milne, A. M., Ed. ***Educational Reform and Vocational Education***. Washington, DC: Office of Educational Research and Improvement and National Assessment of Vocational Education, U.S. Department of Education, 1998. (ERIC Document Reproduction Service No. ED 421 659)

Moss, C. M. "Systematic Self-Reflection: Professional Development for the Reflective Practitioner." Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL, 1997. (ERIC Document Reproduction Service No. ED 411 225)

National Occupational Competency Testing Institute. ***Building a World-class Work Force.*** Big Rapids, MI: NOCTI, n.d.

Neill, D. M. "Transforming Student Assessment." ***Phi Delta Kappan*** 79, no. 1 (September 1997): 34-40, 58. (ERIC No. EJ 550 559)

Newmann, F. M., and Wehlage, G. G. "Five Standards of Authentic Instruction." ***Educational Leadership*** 50, no. 7 (April 1993): 8-12. (ERIC No. EJ 461 121)

Office of Technology Assessment. ***Testing and Assessment in Vocational Education.*** Washington, DC: OTA, U.S. Congress, 1994. (ERIC Document Reproduction Service No. ED 368 908)

O'Neil, J. "Putting Performance Assessment to the Test." ***Educational Leadership*** 49, no. 8 (May 1992): 14-19. (ERIC No. EJ 444 307)

Parkes, J. "Performance Assessment and Student Motivation: Questioning Construct Relevant Variance." Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL, 1997. (ERIC Document Reproduction Service No. ED 411 312)

Parnell, D. ***Why Do I Have to Learn This?*** Waco, TX: CORD Communications, 1995.

Perkins, D., and Salomon, G. "The Science and Art of Transfer." In ***If Minds Matter: A Foreword to the Future, vol. 1,*** edited by A. L. Costa, J. A. Bellanca, and R. Fogarty, pp. 201-209. Palatine, IL: IRI/Skylight Publishing, Inc., 1992. (ERIC Document Reproduction Service No. ED 379 070)

Phillips, D. C., and Soltis, J. F. ***Perspectives on Learning, 3d ed.*** New York: Teachers College Press, 1998.

Phillips, S. E. ***Legal Implications of High-stakes Assessment: What States Should Know.*** Oak Brook, IL: North Central Regional Educational Laboratory, 1993. (ERIC Document Reproduction Service No. ED 370 985)

Popham, W. J. ***Classroom Assessment: What Teachers Need to Know, 2d ed.*** Boston: Allyn and Bacon, 1999.

Reckase, M. D. "Statistical Test Specifications for Performance Assessment: Is This an Oxymoron?" Paper presented at the Annual Meeting of the National Council of Measurement in Education, Chicago, IL, 1997. (ERIC Document Reproduction Service No. ED 410 283)

Resnick, L. B. "Learning in School and out." ***Educational Researcher*** 16, no. 9 (December 1987): 13-20. (ERIC No. EJ 368 309)

References

Riley, K., and Nuttall, D. "Measuring Performance—National Contexts and Local Realities." In ***Measuring Quality: Education Indicators–United Kingdom and International Perspectives***, edited by K. Riley and D. Nuttall, pp. 122-131. London: Falmer Press, 1994.

Roeber, E. "The Technical and Practical Challenges in Developing Innovative Assessment Approaches for Use in Statewide Assessment Programs." ***Contemporary Education*** 69, no. 1 (Fall 1997): 6-10. (ERIC No. EJ 558 968)

Rothman, R. ***Measuring Up: Standards, Assessment and School Reform***. San Francisco: Jossey-Bass Publishers, 1995.

Schell, J. W., and Black, R. S. "Situated Learning: An Inductive Case Study of a Collaborative Learning Experience." ***Journal of Industrial Teacher Education*** 34, no. 4 (Summer 1997): 5-27. (ERIC No. EJ 548 508)

Schell, J. W., and Rojewski, J. W. "Community and Technical College Teaching for Thoughtful Results." ***Community College Journal of Research and Practice*** 19, no. 2 (March-April 1995): 133-150. (ERIC No. EJ 503 545)

Secretary's Commission on Achieving Necessary Skills. *What Work Requires of Schools: A SCANS Report for America 2000*. Washington, DC.: U.S. Department of Labor, 1991. (ERIC Document Reproduction Service No. ED 332 054)

Shavelson, R.; Baxter, G.; and Gao, X. "Sampling Variability of Performance Assessments." ***Journal of Educational Measurement*** 30, no. 3 (Fall 1993): 215-232. (ERIC No. EJ 469 896)

Shepard, L. A. "Why We Need Better Assessments." ***Educational Leadership*** 46, no. 7 (April 1989): 4-5. (ERIC No. EJ 387 134)

Shepard, L. A.; Flexer, R. J.; Hiebert, E. H.; Marion, S. F.; Mayfield, V.; and Weston, T. J. "Effects of Introducing Classroom Performance Assessments on Student Learning." In ***Second Report on Case Study of the Effects of Alternative Assessment in Instruction***. Los Angeles, CA: National Center for Research on Evaluation Standards, and Student Testing, 1994. (ERIC Document Reproduction Service No. ED 379 353)

Simon, K., and Gregg, S. ***Alternative Assessment–Can Real-world Skills Be Tested? Policy Briefs.*** Charleston, WV: Appalachia Educational Laboratory, 1993. (ERIC Document Reproduction Service No. ED 362 575)

Skinner, B. F. ***Science and Human Behavior.*** New York: Macmillian, 1966.

Stecher, B. "The Cost of Performance Assessment in Science: The Rand Perspective." Paper presented at the Annual Meeting of the National Council on Measurement in Education, San Francisco, CA, April 1995. (ERIC Document Reproduction Service No. ED 383 732)

Stecher, B.; Farris, H.; and Hamilton, E. "Performance Standards and Measures." In ***Educational Reform and Vocational Education***, edited by A. M. Milne, pp. 247-264. Washington, DC: Office of Educational Research and Improvement and National Assessment of Vocational Education, U.S. Department of Education, 1998. (ERIC Document Reproduction Service No. ED 421 659)

Stecher, B. M.; Rahn, M. L.; Ruby, A.; Alt, M. N.; Robyn, A.; and Ward, B. ***Using Alternative Assessment in Vocational Education.*** Berkeley: National Center for Research in Vocational Education, University of California, 1997. (ERIC Document Reproduction Service No. ED 400 465)

Sternberg, R. J. and Frensch, P. A. "Mechanisms of Transfer." In ***Transfer on Trial: Intelligence, Cognition, and Instruction***, edited by D. K. Detterman and R. J. Sternberg, pp. 25-38. Norwood, NJ: Ablex, 1993.

Tanner, D. E. "The Long (Suit) and the Short (Comings) of Authentic Assessment." Paper presented at the 49th Annual Meeting of the American Association of Colleges for Teacher Education, Phoenix, AZ, February 26-March 1, 1997. (ERIC Document Reproduction Service No. ED 405 326)

Thorndike, E. L. ***Educational Psychology: The Psychology of Learning***. New York: Teachers College Press, 1913.

Vavrus, L. "Put Portfolios to the Test." ***Instructor*** 100, no. 1 (August 1990): 48-53. (ERIC No. EJ 417 471)

von Glasersfeld, E. ***Radical Constructivism: A Way of Knowing and Learning***. London: Falmer Press, 1995.

Vygotsky, L. S. ***Mind in Society***, edited and translated by M. Cole, V. John-Steiner, S. Scribner, and E. Souberman. Cambridge, MA: Harvard University Press, 1978.

Watson, J. B. ***Behaviorism***. New York: W.W. Norton, 1930.

Weber, E. ***Student Assessment that Works: A Practical Approach***. Boston: Allyn and Bacon, 1999.

Wenger, E. ***Communities of Practice.*** Cambridge, England: Cambridge University Press, 1998.

Wiggins, G. P. ***Assessing Student Performance: Exploring the Purpose and Limits of Testing.*** San Francisco: Jossey-Bass Publishers, 1993.

Wildavsky, B. "Achievement Testing Gets Its Day in Court." ***U.S. News and World Report*** 127, no. 12 (1999): 30-31.

Yen, W. "Scaling Performance Assessments: Strategies for Managing Local Item Dependence." ***Journal of Educational Measurement*** 30, no. 3 (Fall 1993): 187-213. (ERIC No. EJ 469 895)

Using Authentic Assessment in Vocational Education

by Rodney L. Custer
John W. Schell
Brian McAlister
John Scott
Marie Hoepfl

Discusses the theoretical foundations of authentic assessment, examines research on its value and effectiveness in promoting learning. Details strategies and tools for assessing learning and describes federal and state initiatives for using authentic assessment on a large scale.

Order No. IN 381, $9.75

Ordering Information

To order additional copies, please use order number and full title of publication. Orders should be made by school/agency purchase order, letterhead, or prepayment. Personal orders must be prepaid. Make remittance payable to the Center on Education and Training for Employment. Mail order to:

Center on Education and Training for Employment
Center Publications
1900 Kenny Road
Columbus, OH 43210-1090

Phone:	614/292-4277
Toll Free:	800/848-4815, ext. 24277
Fax:	614/292-1260
Web:	http://cete.org/products/index.html

Full payment is due 30 days after receipt of materials and invoice. All payments must be made in U.S. dollars.

Credit Cards
To charge your purchase, indicate either VISA or MasterCard and provide card number, expiration date, amount to be charged, name as it appears on card, telephone number, authorized signature, and the current date.

Quantity Discounts
Orders of 10 or more items, as listed by publication number and title, with a total dollar value for the order of: **$201 to $300, the discount is 5%; $301 to $400, the discount is 10%; $401 and above, the discount is 15%.**

Shipping and Handling
Prices include standard shipping and handling by the best direct method within the continental United States. Priority and international shipments will be charged at actual shipping costs plus a $5 handling fee. Prices listed are in effect at the time of publication of this book. Prices are subject to change without notice.

Center on Education and Training for Employment

THE OHIO STATE UNIVERSITY